Unit A1

Sets, functions and vectors

Introduction to Book A

M208 covers a wide range of pure mathematics, and each book apart from this one concentrates on one topic. This book is different, because it covers the main concepts that underlie the topics in the other books.

In Unit A1 you will review some of the important foundations of pure mathematics and the mathematical language used to describe them. You will start with the plane, and revise ideas relating to points, lines and circles. You will then study in detail the mathematical ideas of a *set* (mostly of numbers or of points in the plane), and a *function*, including functions of real numbers and functions of points in the plane. Finally, you will consider *vectors* in the plane and in three-dimensional space.

In Unit A2 you will look at number systems and their properties. You will first consider *real numbers*, and sets of real numbers, such as the integers and the rational numbers, then study *complex numbers*, investigate their properties, and look at some functions of complex numbers. Finally, you will study *modular arithmetic*, which provides examples of *finite* number systems.

In Unit A3 you will concentrate on mathematical language and communication. You will study the important subject of mathematical proof, including the use of different methods of proof, and how to disprove a statement by finding a counterexample. You will also consider errors in mathematical arguments including errors in deduction. Finally, you will study *equivalence relations* and the idea of a *partition* of a set.

In Unit A4 you will concentrate on *real functions*, and on how to draw their graphs. You will review the graphs of various common functions, and consider a wide range of functions and their properties, including *trigonometric* and *hyperbolic functions*. Finally, you will consider curves that are not the graphs of real functions including *conics* (circles, parabolas, hyperbolas and ellipses) and see that they can be described in terms of a single parameter.

Introduction

In this unit you will look at some of the most fundamental mathematical concepts underlying pure mathematics. Many of these concepts should not be new to you, but working through this unit should ensure that you understand them to the level needed for M208.

Sections 1 to 3 contain basic material that will be crucial throughout the module. It is vital that you become familiar and confident with the ideas and notation introduced in these sections. Section 4 revises concepts that will be used later in the module, in particular in Book C *Linear algebra*.

1 Points, lines and distance

In this section you will revise points, lines and distance in two- and three-dimensional space.

1.1 The plane

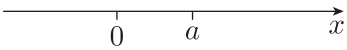


Figure 1 The real line

The set of all real numbers is denoted by $\mathbb{R}$, and this set can be pictured as an infinitely long number line, often called the **real line**, as shown in Figure 1. Each real number a corresponds to a point on the line.

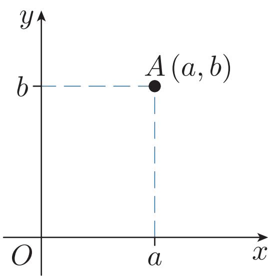


Figure 2 Cartesian coordinates

In this subsection we consider **the plane**, or two-dimensional space. To allow us to specify the locations of points in the plane, we usually use a pair of perpendicular axes, known as **Cartesian** or **rectangular** axes. We usually label the axes x and y; we refer to their intersection point as the **origin** and sometimes label it O. Finally, we choose a unit of distance. The location of any **point** in the plane can be specified by using an ordered pair (a, b) of real numbers, known as **Cartesian coordinates** or just **coordinates**, that give the position of the point relative to the axes, as shown in Figure 2. (An **ordered pair** is a pair in which order matters; for example, the ordered pair $(2, 3)$ is different from the ordered pair $(3, 2)$.) We write $A(a, b)$ to specify the point A with coordinates (a, b).

It is important to understand that the coordinates of a point depend on where the axes have been placed in the plane; if we had chosen the axes to be in a different position, then usually the coordinates of the point would be different. However, once we have chosen the position of the axes, we often do not bother to distinguish explicitly between a point and its representation using these coordinates: we simply write (a, b) to denote the point A.

We use the notation $\mathbb{R}^2$ to denote the plane.

René Descartes

The adjective Cartesian comes from the surname of the French mathematician and philosopher René Descartes (1596–1650). He was the first person to show in print how algebra could be used to study geometry, in his 1637 publication *La géométrie*. Descartes' procedure differed from the system of Cartesian coordinates that we use today. His axes were not necessarily at right angles, and could be chosen in relation to the circumstances of the problem rather than being given in advance.

The plane, together with an **origin** O and a pair of x- and y-axes, is known as **two-dimensional Euclidean space**.

Euclidean space is named after the Greek mathematician Euclid. Little is known for certain about Euclid but he is believed to have worked in Alexandria in around 300 BCE.

Euclid's *Elements*, a mathematical treatise of thirteen books which had its origins on papyrus rolls, has become one of the most frequently printed texts of all time. Although *Elements* covers both plane and solid Euclidean geometry, Euclid had no notion of axes or coordinates.

Lines

The equation of any straight line in $\mathbb{R}^2$, except a line parallel to the y-axis, can be written in the form

$$y = mx + c,$$

where $m, c \in \mathbb{R}$.

In this equation:

- m is the **gradient** (or *slope*) of the line, given by
$$m = \frac{y_2 - y_1}{x_2 - x_1}, \tag{1}$$
where (x_1, y_1) and (x_2, y_2) are any two points on the line such that $x_1 \neq x_2$
- c is the **y-intercept** of the line; that is, $(0, c)$ is the point at which the line crosses the y-axis, as illustrated in Figure 3(a).

The line with gradient m that crosses the y-axis at the origin has equation $y = mx$, since $c = 0$ in this case; see Figure 3(b). The horizontal line (parallel to the x-axis) with y-intercept c has equation $y = c$, since the gradient $m = 0$ in this case; see Figure 3(c).

The equation of a line parallel to the y-axis cannot be written in the form (1). The vertical line (parallel to the y-axis) with x-intercept a has equation $x = a$; see Figure 3(d). The equation of such a line cannot be written in the form $y = mx + c$ because the gradient is undefined.

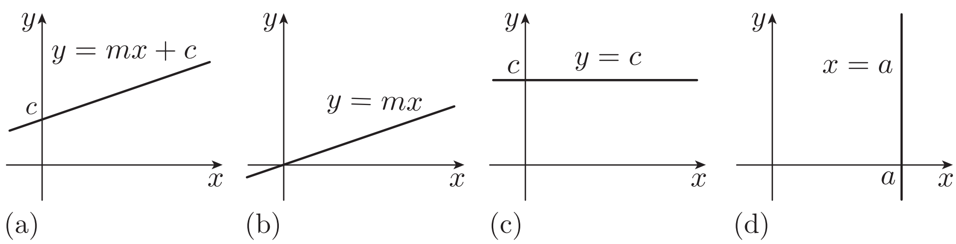


Figure 3 Lines in the plane

In all of the cases above, the equation of the line in the plane can be rearranged to take the form

$$ax + by = c, \tag{2}$$

for some real numbers a, b and c, where a and b are not both zero. (Note that the numbers a and c here are not the same as those called a and c in Figure 3.)

In fact, any line in $\mathbb{R}^2$ has an equation of the form (2) and, conversely, any equation of the form (2) represents a line in $\mathbb{R}^2$.

Equation of a line

The general equation of a line in $\mathbb{R}^2$ is

$$ax + by = c,$$

where a, b and c are real numbers, and a and b are not both zero.

From formula (1) for the gradient of a line, we can see that the equation of the line with gradient m that passes through the point (x_1, y_1) is

$$y - y_1 = m(x - x_1).$$

Exercise A1

Determine the equation of the line with gradient -3 that passes through the point $(2, -1)$.

Exercise A2

Determine the equation of the line through each of the following pairs of points.

(a) $(1, 1)$ and $(3, 5)$ (b) $(0, 0)$ and $(0, 8)$ (c) $(0, 0)$ and $(4, 2)$

(d) $(4, -1)$ and $(2, -1)$

Parallel and perpendicular lines

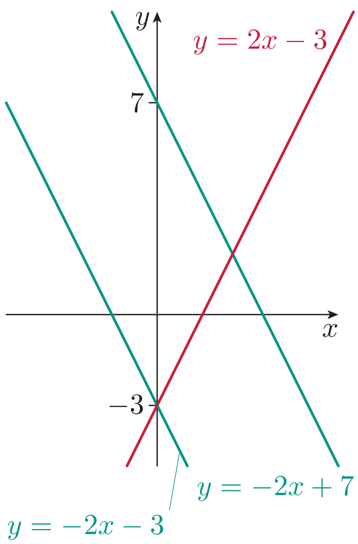


Figure 4 Parallel and perpendicular lines

Two distinct lines are **parallel** if they never meet, and **perpendicular** if they meet at right angles.

Saying that two non-vertical lines are parallel is equivalent to saying that they have the same gradient but different y-intercepts. For example, as shown in Figure 4, the lines $y = -2x + 7$ and $y = -2x - 3$ are parallel since they both have gradient -2 but their y-intercepts are 7 and -3, respectively, whereas the lines $y = -2x + 7$ and $y = 2x - 3$ are not parallel since their gradients -2 and 2 are not equal.

We can also use the gradients of a pair of non-vertical lines to check whether they are perpendicular, as follows.

Gradients of perpendicular lines

Let l_1 and l_2 be lines with gradients m_1 and m_2, respectively.

- If l_1 and l_2 are perpendicular, then $m_1 m_2 = -1$.
- If $m_1 m_2 = -1$, then l_1 and l_2 are perpendicular.

To see that the first statement in the box is true, suppose that the lines l_1 and l_2 are perpendicular and that neither line is vertical. Let the gradients of l_1 and l_2 be m_1 and m_2, respectively. Then one of the lines (l_1, say) must slope up from left to right and the other (l_2, say) must slope down from left to right, as shown in Figure 5.

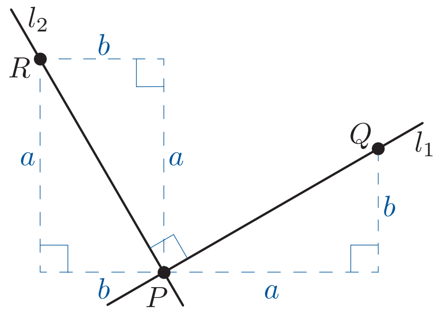


Figure 5 Perpendicular lines

Let the lines intersect at P, and let Q be a point on l_1 to the right of P. Suppose that Q is a units to the right of P and b units up from P, as illustrated in Figure 5. Let R be the point on l_2 obtained by rotating PQ anticlockwise through a right angle; then R is b units to the left of P and a units up from P, as shown.

It follows that the gradient of l_1 is $m_1 = b/a$, and the gradient of l_2 is $m_2 = -a/b$. Hence

$$m_1 m_2 = \frac{b}{a} \times \left(-\frac{a}{b}\right) = -1.$$

The proof of the second statement in the box above is not given here.

Worked Exercise A1

Determine which of the following lines are parallel, and which are perpendicular to each other.

$l_1\colon\ y = -2x + 4 \qquad l_2\colon\ 2x - 3y - 2 = 0 \qquad l_3\colon\ y - 2x = 9$

$l_4\colon\ 2y + 3x + 5 = 0 \qquad l_5\colon\ x + \frac{1}{2}y + 2 = 0 \qquad l_6\colon\ 2y = 3x + 7$

Solution

We can rearrange the equations of the lines to find their gradients as follows.

$l_1\colon\ y = -2x + 4 \qquad l_2\colon\ y = \frac{2}{3}x - \frac{2}{3} \qquad l_3\colon\ y = 2x + 9$

$l_4\colon\ y = -\frac{3}{2}x - \frac{5}{2} \qquad l_5\colon\ y = -2x - 4 \qquad l_6\colon\ y = \frac{3}{2}x + \frac{7}{2}$

Thus the gradients of the lines are -2, $\frac{2}{3}$, 2, $-\frac{3}{2}$, -2 and $\frac{3}{2}$, respectively.

The gradients of l_1 and l_5 are equal, and since their y-intercepts are different, the lines l_1 and l_5 are parallel. Multiplying the gradients of l_2 and l_4 gives $\frac{2}{3} \times \left(-\frac{3}{2}\right) = -1$, so the lines l_2 and l_4 are perpendicular.

Exercise A3

Determine which of the following lines are parallel, and which are perpendicular to each other.

l_1: $y = -2x + 4$ l_2: $6x - 3y + 4 = 0$ l_3: $2y + x = 10$

l_4: $6y - 3x + 5 = 0$ l_5: $x - 2y + 2 = 0$ l_6: $2y + 4x + 7 = 0$

Distance between two points in the plane

Next, we find the formula for the distance between any two points in the plane.

We use the idea of the **modulus** of a real number k, written $|k|$ and defined by

$$|k| = \begin{cases} k, & \text{if } k \geq 0, \\ -k, & \text{if } k < 0. \end{cases}$$

(The modulus of k, usually read as 'mod k' is sometimes called the **absolute value** or **magnitude** of k.)

Suppose that $P(x_1, y_1)$ and $Q(x_2, y_2)$ are two points in the plane, as shown in Figure 6. We can construct a right-angled triangle PNQ as shown: the line PN is parallel to the x-axis, the line QN is parallel to the y-axis, the angle PNQ is a right angle, and PQ is the hypotenuse of the triangle. In Figure 6, P and Q are drawn in the first quadrant and with PQ sloping up from left to right, but the formula holds wherever the points are in the plane.

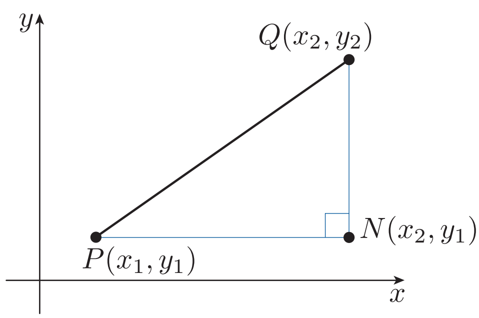


Figure 6 Distance between P and Q in the plane

The length of PN is $|x_2 - x_1|$ and the length of QN is $|y_2 - y_1|$. It follows from Pythagoras' Theorem that

$$PQ^2 = PN^2 + QN^2,$$

and since $|k|^2 = k^2$ for any real number k, we have

$$PQ = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Distance formula for $\mathbb{R}^2$

The distance between the two points (x_1, y_1) and (x_2, y_2) in the plane is

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

For example, it follows from the formula above that the distance between the points $(1, 2)$ and $(3, -4)$ is

$$\begin{aligned}\sqrt{(3-1)^2 + (-4-2)^2} &= \sqrt{2^2 + (-6)^2} \\ &= \sqrt{40} = \sqrt{4 \times 10} \\ &= \sqrt{4}\sqrt{10} = 2\sqrt{10}.\end{aligned}$$

Exercise A4

Find the distances between the following pairs of points in the plane.

(a) $(0, 0)$ and $(5, 0)$ (b) $(0, 0)$ and $(3, 4)$ (c) $(1, 2)$ and $(5, 1)$

(d) $(3, -8)$ and $(-1, 4)$

Circles

A **circle** in $\mathbb{R}^2$, as illustrated in Figure 7, is the set of points $P(x, y)$ that lie at a fixed distance r, called the **radius**, from a fixed point $C(a, b)$, called the **centre** of the circle.

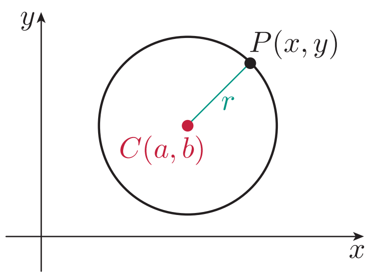


Figure 7 A circle with radius r and centre (a, b)

By the distance formula, every point (x, y) on the circle with centre (a, b) and radius r satisfies the equation

$$\sqrt{(x-a)^2 + (y-b)^2} = r.$$

Squaring this equation to remove the square root gives the following.

Equation of a circle

The equation of the circle in $\mathbb{R}^2$ with centre (a, b) and radius r is

$$(x-a)^2 + (y-b)^2 = r^2.$$

In this unit we will just work with equations of circles in this form, without multiplying out the brackets. In Unit A4 *Real functions, graphs and conics*, you will see how multiplying out the brackets leads to other forms for the equations of circles.

Worked Exercise A2

Find the equation of the circle with centre $(-1, 2)$ and radius $\sqrt{3}$.

Solution

The circle has equation

$$(x-(-1))^2+(y-2)^2=(\sqrt{3})^2,$$

that is,

$$(x+1)^2+(y-2)^2=3.$$

Exercise A5

Determine the equation of each of the following circles, given the centre and radius.

(a) Centre the origin, radius 4.

(b) Centre $(-1, 0)$, radius $\sqrt{2}$.

(c) Centre $(3, -4)$, radius 2.

1.2 Three-dimensional space

We now look briefly at three-dimensional space.

We define a coordinate system in three-dimensional space using three *mutually perpendicular* axes. The word *mutually* here means that the condition holds for any pair, so **mutually perpendicular** means that any two of the axes are perpendicular.

First, we choose a point O as the origin, and then we choose an x-axis and a y-axis at right angles to each other. Next, we draw a third line through the origin, perpendicular both to the x-axis and to the y-axis; this line is called the z-axis. We choose the positive direction of the z-axis to be such that the x-, y- and z-axes form a so-called **right-handed system of axes**. This means that if you hold the thumb and first and second fingers of your right hand at right angles to each other, and label them x, y and z, in that order, then you can turn your hand in such a way that your fingers point in the positive directions of the corresponding axes, as shown in Figure 8.

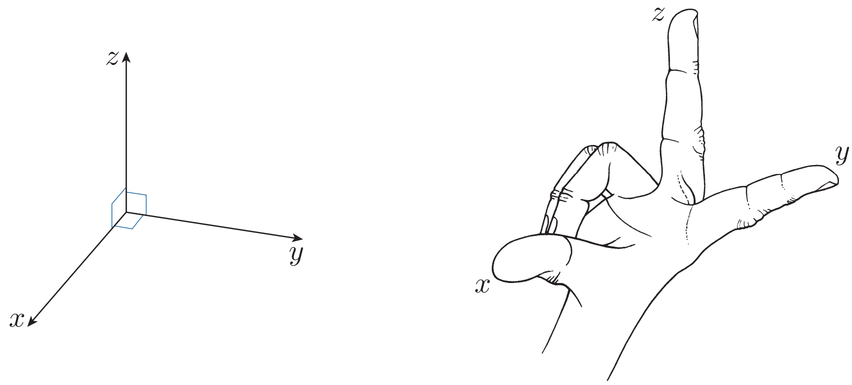


Figure 8 A right-handed system of coordinate axes for $\mathbb{R}^3$

Finally, we choose a unit of distance.

We represent each point in three-dimensional space by an **ordered triple** (a, b, c) of real numbers. The point with coordinates (a, b, c) is reached from the origin by moving a distance a in the direction of the x-axis, a distance b in the direction of the y-axis, and a distance c in the direction of the z-axis, as illustrated in Figure 9(a).

For instance, the point with coordinates $(-3, -2, 4)$ is shown in Figure 9(b).

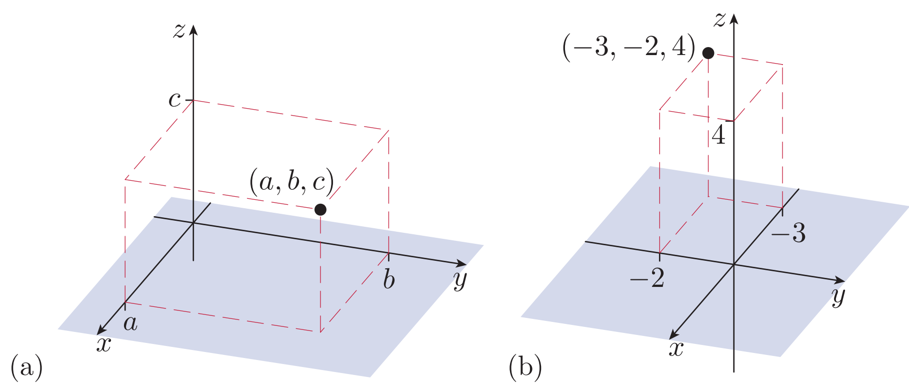


Figure 9 Three-dimensional Cartesian coordinates

In Figure 9, the plane containing the x-axis and the y-axis is shaded. Usually we think of this plane as being horizontal, and the z-axis as being vertical.

We use the notation $\mathbb{R}^3$ to denote three-dimensional space.

Exercise A6

Sketch the x-, y- and z-axes and the points with coordinates $(0, 1, 2)$ and $(-1, 2, -1)$.

As with $\mathbb{R}^2$, once we have chosen the position of the axes, we often do not bother to distinguish explicitly between a point and its representation using these coordinates; we simply write (a, b, c) to denote the point in $\mathbb{R}^3$ represented by this triple.

Three-dimensional space, together with an origin and a set of x-, y- and z-axes, is known as **three-dimensional Euclidean space**.

Distance between points in $\mathbb{R}^3$

You saw in Subsection 1.1 that the distance between two points (x_1, y_1) and (x_2, y_2) in the plane is given by

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

We can establish a similar formula for the distance between two points in $\mathbb{R}^3$, as follows.

Let $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$ be two points in $\mathbb{R}^3$. Let M be the point (x_2, y_2, z_1); then M lies in the same horizontal plane as P, and MQ is parallel to the z-axis. Next, let N be the point (x_1, y_2, z_1); then N also lies in the same horizontal plane as P, and MN and NP are parallel to the x- and y-axes, respectively.

The triangles PQM and PMN are both right-angled triangles, with right angles at M and N, respectively, as shown in Figure 10.

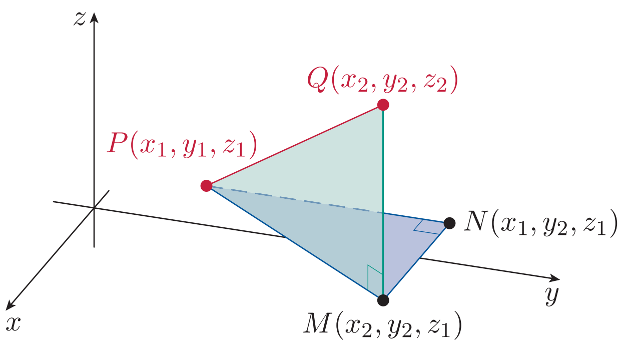


Figure 10 Distance between P and Q in $\mathbb{R}^3$

The length of PN is $|y_2 - y_1|$ and the length of NM is $|x_2 - x_1|$. It follows from Pythagoras' Theorem that

$$PM^2 = NM^2 + PN^2,$$

so

$$PM^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2.$$

Using Pythagoras' Theorem again gives

$$PQ^2 = PM^2 + MQ^2,$$

and since the length of MQ is $|z_2 - z_1|$ we obtain

$$PQ^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2,$$

that is,

$$PQ = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Distance formula for $\mathbb{R}^3$

The distance between the two points (x_1, y_1, z_1) and (x_2, y_2, z_2) in $\mathbb{R}^3$ is

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

For example, it follows from this formula that the distance between the points $(1, 2, 3)$ and $(4, -2, 15)$ is

$$\sqrt{(4-1)^2 + (-2-2)^2 + (15-3)^2} = \sqrt{169} = 13.$$

Exercise A7

Find the distances between the following pairs of points in $\mathbb{R}^3$.

(a) $(1, 1, 1)$ and $(4, 1, -3)$ (b) $(1, 2, 3)$ and $(3, 0, 3)$

We will return to the topic of three-dimensional space in Section 4, where we will consider vectors in $\mathbb{R}^3$ as well as in $\mathbb{R}^2$, and find the general equation of a plane in $\mathbb{R}^3$.

2 Sets

In this section you will revise the notion of sets, learn new notation for describing sets, and practise working with sets and set notation. These skills will be crucial in the rest of the module.

2.1 What is a set?

In mathematics we frequently work with collections of objects of various kinds. We may, for example, consider the following:

- solutions of a quadratic equation
- points on a circle
- vertices of a triangle
- points on a plane in $\mathbb{R}^3$
- even numbers less than 100
- students taking a particular examination.

The concept of a *set* allows us to work with such collections systematically.

You can think of a **set** as a collection of objects, such as numbers, points, functions, or even a collection of other sets. Each object in a set is an **element** or **member** of the set, and the elements *belong to* the set, or are *in* the set.

There is no restriction on the types of object that may appear in a set, provided that the set is specified in a way that enables us to decide, in principle, whether a given object is in the set.

There are many ways of making such a specification. For example, we can define S to be the set of numbers in the list

$$4,\ 9,\ 3,\ 2.$$

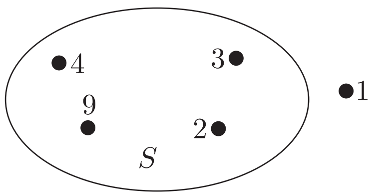


Figure 11 A Venn diagram of the set S

This enables us to decide that the number 2 (say) is in S, but that the number 1 (say) is not in S. We can illustrate this set by a diagram, as in Figure 11, where the symbol S is *not* a member of the set but a label for it. (Similar labels will appear in other diagrams.) Such a diagram is called a **Venn diagram**, after the nineteenth-century Cambridge mathematician John Venn.

We can also define a set by describing its elements; for example,

let E be the set of all even integers.

This description enables us to determine whether a given object is in E by deciding whether it is an even integer; for example, 6 is in E, but 5 is not.

Some sets are used so often that special symbols are reserved for them. Recall that a **real number** is a number with a decimal expansion (possibly infinite), for example, 1.1 or $\pi = 3.14\ldots$, and a **rational number** is a real number that can be expressed as a fraction, for example, $14/5$ or $-3/4$. You will revise these sets more thoroughly in Unit A2 *Number systems*. We use the following notation, some of which you met in Section 1.

$\mathbb{R}$ denotes the set of real numbers.

$\mathbb{R}^*$ denotes the set of non-zero real numbers.

$\mathbb{Q}$ denotes the set of rational numbers.

$\mathbb{Z}$ denotes the set of integers $\ldots, -2, -1, 0, 1, 2, \ldots$.

$\mathbb{N}$ denotes the set of natural numbers $1, 2, 3, \ldots$.

A **finite set** is a set that has a finite number of elements; that is, the number of elements is some natural number, or 0. Any set that is not a finite set is an **infinite set**.

We use the symbol $\in$ to indicate membership of a set; for example, we indicate that 7 is a member of $\mathbb{N}$ by writing

$7 \in \mathbb{N}$. (This is usually read as '7 belongs to $\mathbb{N}$' or '7 is in $\mathbb{N}$'.)

We indicate that -9 is *not* a member of $\mathbb{N}$ by writing

$-9 \notin \mathbb{N}$. ('-9 does not belong to $\mathbb{N}$' or '-9 is not in $\mathbb{N}$'.)

We also use the symbol $\in$ when we wish to introduce a symbol that stands for an *arbitrary* (that is, general, unspecified) element of a set. For example, to indicate that x is a **real variable**, that is, an arbitrary member of the set $\mathbb{R}$, we write

let $x \in \mathbb{R}$.

We often write $x_1, x_2 \in S$ as shorthand to combine $x_1 \in S$ and $x_2 \in S$.

Exercise A8

Which of the following statements are true?

(a) $-3 \in \mathbb{Z}$ (b) $5 \notin \mathbb{N}$ (c) $1.3 \notin \mathbb{Q}$ (d) $1, 3 \in \mathbb{Q}$

(e) $-\pi \in \mathbb{R}$ (f) $\frac{1}{2} \in \mathbb{N}$ (g) $0, 1 \in \mathbb{R}^*$ (h) $\sqrt{2} \notin \mathbb{R}$

2.2 Set notation

We now look at some formal ways of specifying a set.

We can specify a set with a small number of elements by listing these elements between a pair of braces (curly brackets). For example, we can specify the set A consisting of the first five natural numbers, illustrated in Figure 12, by

$$A = \{1, 2, 3, 4, 5\}.$$

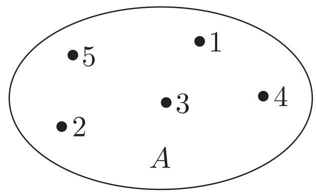


Figure 12 The set A

The membership of a set is not affected by the order in which its elements are listed, so we can specify this set A equally well by

$$A = \{5, 2, 1, 4, 3\}.$$

Similarly, we can specify the set B of vertices of the square shown in Figure 13 by

$$B = \{(0, 0), (1, 0), (1, 1), (0, 1)\}.$$

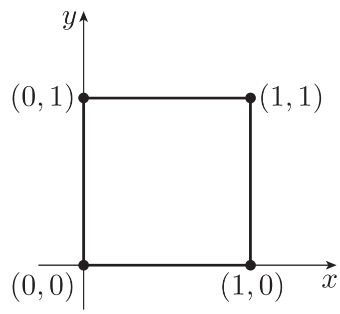


Figure 13 The set B

We can even specify a set C, illustrated in Figure 14, whose elements are the three sets $\{1, 3, 5\}$, $\{9, 4\}$ and $\{2\}$ by

$$C = \{\{1, 3, 5\}, \{9, 4\}, \{2\}\}.$$

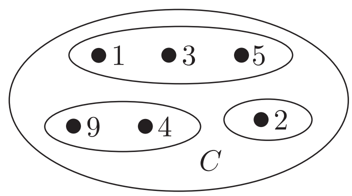


Figure 14 The set C

A set with only one element, such as the set $\{2\}$, is called a **singleton** or a **singleton set**. (Do not confuse the *set* $\{2\}$ which contains the number 2, with the *number* 2 itself.)

Exercise A9

Which of the following statements are true?

(a) $1 \in \{4, 3, 1, 7\}$

(b) $\{-9\} \in \{\{6, 1, 2\}, \{8, 7, 9, 5\}, \{-9\}, \{5, 4\}\}$

(c) $\{9\} \in \{5, 6, 7, 8, 9\}$

(d) $(0, 1) \in \{(1, 0), (1, 4), (2, 4)\}$

(e) $1, 0 \in \{(1, 0), (1, 4), (2, 4)\}$

(f) $\{1, 0\} \in \{\{0, 1\}, \{1, 4\}, \{2, 4\}\}$

It does not matter if we specify a set element more than once within set brackets. For example,

$$\{1, 2, 3, 3\} \quad \text{and} \quad \{1, 2, 3\}$$

describe the same set. However, we usually try to avoid specifying an element more than once.

For a set with a large number of elements, it is not practical to list all the elements, so we sometimes use three dots (called an **ellipsis**) to indicate that a particular pattern of membership continues. For example, we can specify the set consisting of the first 100 natural numbers by writing

$$\{1, 2, 3, \ldots, 100\}.$$

The use of an ellipsis can be extended to certain infinite sets. For example, we can specify the set of all natural numbers by writing

$$\{1, 2, 3, \ldots\}.$$

One disadvantage of this notation is that the pattern indicated by the ellipsis may be ambiguous. For example, it is not clear whether

$$\{3, 5, 7, \ldots\}$$

denotes the set of odd prime numbers or the set of odd natural numbers greater than 1. For this reason, this notation can be used only when the pattern of membership is obvious, or where an additional clarifying explanation is given.

An alternative way of specifying a set is to use variables to build up objects of the required type, and then write down the condition(s) that the variables must satisfy. For example, consider the set of all real numbers x such that $x > 3$. Using set notation, we write this as

$$\{x \in \mathbb{R} : x > 3\},$$

which is read as shown in Figure 15.

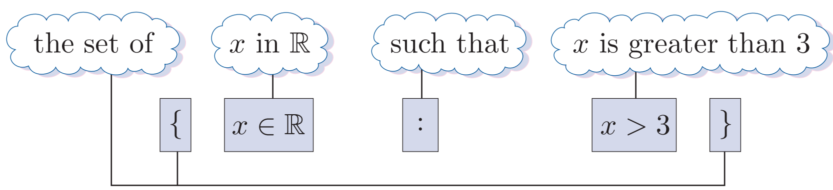


Figure 15 How to ‘read’ set notation

A set can often be described in several different ways using such set notation. In particular, we can use a letter other than x to denote an arbitrary (general) element of a set; for example, the set above can also be written as

$$\{r \in \mathbb{R} : r > 3\}.$$

If it is necessary to include more than one condition after the colon, then we write either a comma or the word ‘and’ between the conditions. So the

set of real numbers greater than 0, and less than or equal to 1, can be written as

$$\{x \in \mathbb{R} : x > 0,\ x \leq 1\} \quad \text{or} \quad \{x \in \mathbb{R} : x > 0 \text{ and } x \leq 1\},$$

although usually we combine the inequalities and write

$$\{x \in \mathbb{R} : 0 < x \leq 1\}.$$

Sometimes it is convenient to specify a set by writing an expression in one or more variables before the colon, and the conditions on the variables after the colon. For example, the set of even integers less than 100 may be specified by

$$\{2k : k \in \mathbb{Z} \text{ and } k < 50\}.$$

Just as when listing the elements of a set, it does not matter when using set notation if a set element is specified more than once. For example,

$$\{\sin x : x \in \mathbb{R}\}$$

specifies the same set as

$$\{\sin x : 0 \leq x < 2\pi\}.$$

Exercise A10

Which of the following statements are true?

(a) $\frac{9}{2} \in \{x \in \mathbb{R} : x > 3\}$ (b) $7 \in \{3k+1 : k \in \mathbb{Z}\}$

(c) $-\frac{7}{2} \in \{x \in \mathbb{Z} : x < 5\}$ (d) $8 \in \{2^x : x \in \mathbb{R},\ 0 < x < 2\}$

(e) $9 \in \{n \in \mathbb{Z} : n = k^2 \text{ for some } k \in \mathbb{Z}\}$ (f) $6 \in \{m(m-1) : m \in \mathbb{N}\}$

(g) $4 \in \{r : r \text{ is an even integer},\ 0 < r < 4\}$

Notice that the next worked exercise contains lines of blue text, marked with the icons . You will see similar text in some of the worked exercises and proofs throughout this module. This text tells you what someone doing the mathematics might be thinking, but would not write down; or what a lecturer might say to explain the thinking behind the mathematics, but would not write on the board. It should help you understand how you might approach a similar exercise yourself.

Worked Exercise A3

Use set notation to specify each of the following.

(a) The set of all natural numbers greater than 50.

(b) The set of all odd integers.

Solution

(a) The elements of this set are the natural numbers n such that $n > 50$.

The set is $\{n \in \mathbb{N} : n > 50\}$.

(b) The odd integers are the numbers that can be written in the form $2k + 1$, for some integer k.

The set is $\{2k + 1 : k \in \mathbb{Z}\}$.

The choice of the variables is arbitrary in these sets, but k for an integer and n for a natural number are conventional.

Exercise A11

Use set notation to specify each of the following.

(a) The set of integers greater than -2 and less than 1000.

(b) The set of positive rational numbers with square greater than 2.

(c) The set of even natural numbers.

(d) The set of integer powers of 2.

Set notation is useful when we wish to refer to the set of solutions of one or more equations (called the **solution set**). For example, the real solutions of the equation $x^2 = 1$ form the set

$$\{x \in \mathbb{R} : x^2 = 1\} = \{-1, 1\}.$$

The solution set of an equation depends on the set of values from which the solutions are taken. For example, the solution set of the equation

$$(x - 1)(2x - 1) = 0$$

is

$$\{x \in \mathbb{R} : (x - 1)(2x - 1) = 0\} = \left\{1, \tfrac{1}{2}\right\}$$

if we are interested in real solutions. However, the solution set is

$$\{x \in \mathbb{Z} : (x - 1)(2x - 1) = 0\} = \{1\}$$

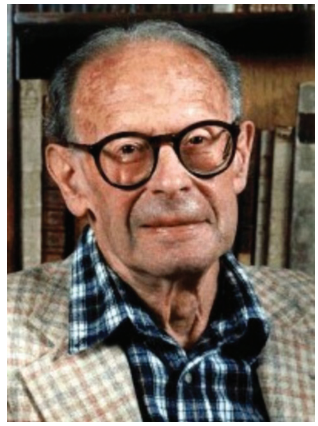
if we are interested only in integer solutions. In this unit we assume that solutions are taken from $\mathbb{R}$ unless otherwise stated.

Sometimes an equation has *no* real solutions, so its solution set has no elements.

The set with no elements arises frequently in mathematics, so it is given a special name and notation. It is called the **empty set** and is denoted by the symbol $\varnothing$. Thus, for example,

$$\{x \in \mathbb{R} : x^2 = -1\} = \varnothing.$$

The symbol for the empty set, $\varnothing$, was introduced in 1939 by the French mathematician André Weil (1906–1998), who took the symbol from the Norwegian alphabet.



André Weil

2.3 Intervals

You saw in Subsection 1.1 that the set of real numbers $\mathbb{R}$ can be pictured as a number line, called the real line. Many sets involve ranges of real numbers extending along the real line from one number a to another number b. Each of the endpoints a and b may be either included or excluded. Such sets are called **intervals** of the real line, and they occur so frequently that we use special notation for them. For example:

- the interval given by $-2 < x < 5$, in which both endpoints are *excluded*, is denoted by $(-2, 5)$ and is an example of an *open* interval
- the interval given by $-2 \leq x \leq 5$, in which both endpoints are *included*, is denoted by $[-2, 5]$ and is an example of a *closed* interval
- the intervals given by $-2 < x \leq 5$ and $-2 \leq x < 5$, in which one endpoint is included and the other is excluded, are denoted by $(-2, 5]$ and $[-2, 5)$, respectively, and are examples of *half-open* (or *half-closed*) intervals.

In some texts, a reversed square bracket is used instead of a round bracket to indicate an excluded endpoint; for example $]-2, 5[$ is used instead of $(-2, 5)$ for an open interval.

We use the symbol ∞ (infinity) when an interval extends indefinitely far to the right on the real line, and the symbol $-\infty$ when an interval extends indefinitely far to the left. For example:

- the set of all real numbers greater than -3 is denoted by $(-3, \infty)$
- the set of all real numbers less than or equal to 4 is denoted by $(-\infty, 4]$.

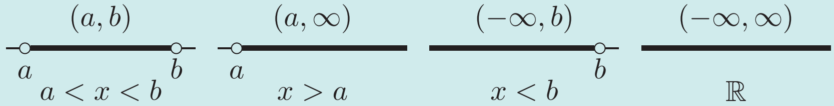
The symbol ∞ does not denote a real number: instead, it simply means that the interval continues indefinitely. We always use round brackets with ∞ and $-\infty$.

The notation for intervals is summarised in the box below.

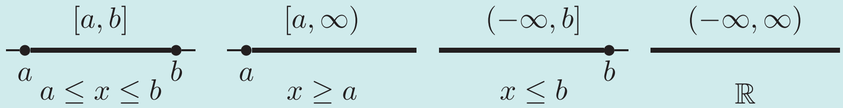
Interval notation

Intervals are denoted as follows.

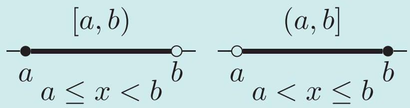
Open intervals



Closed intervals



Half-open (or half-closed) intervals



Remarks

1. In the box above, a hollow dot $\circ$ indicates that an endpoint is excluded, and a solid dot $\bullet$ indicates that an endpoint is included.
2. A singleton set $\{a\}$, containing a single number a, is a closed interval whose endpoints are equal.
3. An interval such as $[a, \infty)$ is regarded as closed, rather than half-open (or half-closed), because it contains *all* the real numbers greater than or equal to a. However, the interval $\mathbb{R} = (-\infty, \infty)$ is considered to be both open and closed.
4. We also use the notation (a, b) to denote a point in the plane, but in most cases it should be obvious whether a point or an interval is intended.

Exercise A12

Which of the following statements are true?

(a) $1 \in (1, 5)$ (b) $1 \in (-1, 1]$ (c) $\infty \in (0, \infty)$ (d) $0 \notin \mathbb{R}^*$

(e) If $x \in \mathbb{R}^*$, then $x \in (0, \infty)$.

Exercise A13

Use interval notation to specify the following intervals.

(a) 

(b) The set of real numbers x such that $-6.5 < x \le 21$.

(c) $\{x \in \mathbb{R} : x > -273\}$.

2.4 Plane sets

In Subsection 1.1 you met the plane $\mathbb{R}^2$, and saw that each point in the plane can be represented as an ordered pair (x, y) with respect to a chosen pair of axes. A set of points in $\mathbb{R}^2$ is called a **plane set** or a **plane figure**. The lines and circles that you met in Subsection 1.1 are simple examples of plane sets.

Lines as plane sets

Consider a straight line l_1 with gradient m and y-intercept c, as illustrated in Figure 16. This line is the set of all points (x, y) in the plane such that $y = mx + c$. Using set notation, we write this as

$$l_1 = \{(x, y) \in \mathbb{R}^2 : y = mx + c\}.$$

(We often refer to 'the line $y = mx + c$' as a shorthand way of specifying this set.)

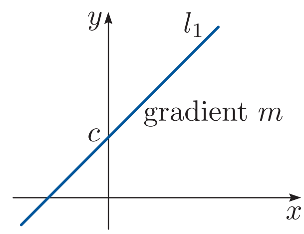


Figure 16 The line l_1

For a line l_2 parallel to the y-axis with x-intercept a, as illustrated in Figure 17, we write

$$l_2 = \{(x, y) \in \mathbb{R}^2 : x = a\}.$$

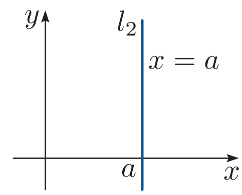


Figure 17 The vertical line l_2

An alternative way of specifying a line is to write an expression for one or both of the coordinates. For example, an alternative way of specifying the line l_1 with equation $y = mx + c$ is

$$l_1 = \{(x, mx + c) : x \in \mathbb{R}\}.$$

It does not matter what variable we use to specify the line. For example, we can also write

$$l_1 = \{(t, mt + c) : t \in \mathbb{R}\}.$$

Exercise A14

(a) Use set notation to specify the line l with gradient 2 that passes through the point $(0, 5)$.

(b) Sketch the line $l = \{(x, y) \in \mathbb{R}^2 : y = 1 - x\}$.

(c) Sketch the line $l = \{(x, x) : x \in \mathbb{R}\}$.

Circles as plane sets

Consider a circle C with centre (a, b) and radius r, as illustrated in Figure 18. This circle is the set of all points (x, y) in the plane such that $(x - a)^2 + (y - b)^2 = r^2$, so, in set notation, it can be written as

$$C = \{(x, y) \in \mathbb{R}^2 : (x - a)^2 + (y - b)^2 = r^2\}.$$

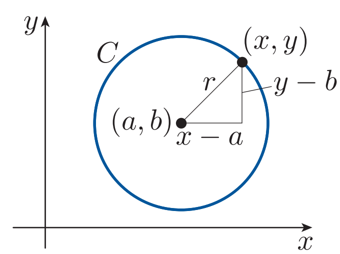


Figure 18 A circle C

The **unit circle** U is defined to be the circle centred at the origin with radius 1, so it is the set of points (x, y) in the plane whose distance from

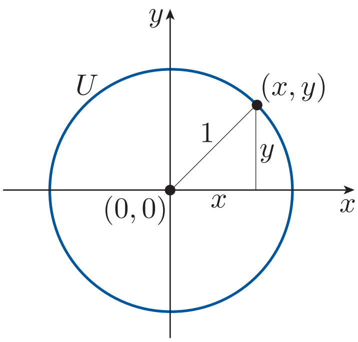


Figure 19 The unit circle U

the origin $(0, 0)$ is 1 (see Figure 19). In set notation, the unit circle can be written as

$$U = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

Exercise A15

(a) Use set notation to specify the circle C of radius 3 centred at $(1, -4)$.

(b) Sketch the circle $C = \{(x, y) \in \mathbb{R}^2 : (x-1)^2 + (y-3)^2 = 4\}$.

Half-planes, discs and other plane sets

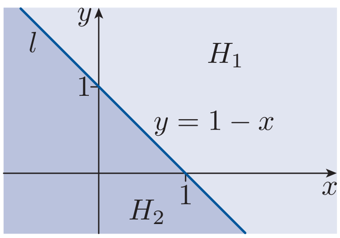


Figure 20 The plane split into three parts by l

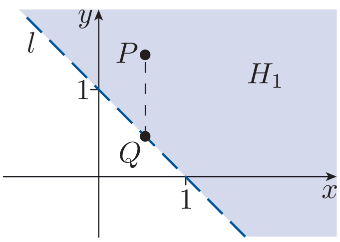


Figure 21 A point P in H_1

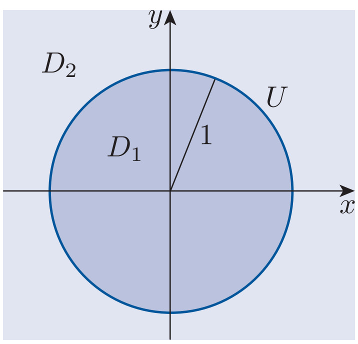


Figure 22 The plane split into three parts by the unit circle U

Consider the line

$$l = \{(x, y) \in \mathbb{R}^2 : y = 1 - x\}.$$

This line splits $\mathbb{R}^2$ into three separate parts, as shown in Figure 20: the line l itself, the set H_1 of points lying *above* the line, and the set H_2 of points lying *below* the line.

For any point $P = (x, y)$ in H_1, the point $Q = (x, 1-x)$ lies on the line l, directly below P, as illustrated in Figure 21, so $y > 1 - x$. Similarly, each point (x, y) in H_2 satisfies $y < 1 - x$. Thus

$$H_1 = \{(x, y) \in \mathbb{R}^2 : y > 1 - x\}$$

and

$$H_2 = \{(x, y) \in \mathbb{R}^2 : y < 1 - x\}.$$

The set of points on one side of a line, possibly together with all the points on the line itself, is known as a **half-plane**. A half-plane that does not include the points on the line can be specified using set notation as in the examples H_1 and H_2 above. The corresponding half-plane that includes the points on the line can be specified by changing the symbol $>$ to $\geq$, or the symbol $<$ to $\leq$.

When we sketch a plane set that *excludes* a boundary line, as for the set H_1 in Figure 21, we draw the boundary as a *broken* line; if the plane set *includes* a boundary line, then we draw the boundary as a *solid* line.

We can treat other plane sets in a similar way. For example, consider the unit circle

$$U = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

This circle splits $\mathbb{R}^2$ into three separate parts, as illustrated in Figure 22: the circle U itself, the set D_1 of points lying *inside* the circle and the set D_2 of points lying *outside* the circle.

The condition for a point (x, y) to lie inside U is that the distance of the point from the origin is less than 1. It follows that the square of the distance of the point (x, y) from the origin is also less than 1, so

$$D_1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}.$$

Similarly,

$$D_2 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 > 1\}.$$

The set of points inside a circle, possibly together with all the points on the circle, is known as a **disc**. Figure 23 shows the disc D_1 with the broken line indicating that the points on the circle are not included in the set.

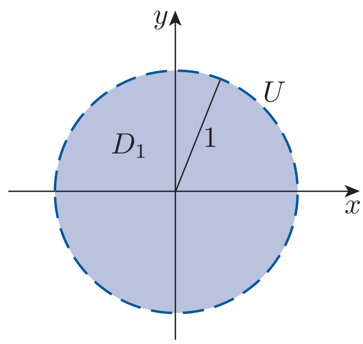


Figure 23 The disc D_1

If we wish to specify the disc consisting of the unit circle together with the points inside it, we replace the inequality $<$ by $\leq$ in the set notation specification of D_1 given above, and draw the boundary as a solid line.

As another example, consider the set of points lying inside the square with vertices $(0, 0)$, $(1, 0)$, $(1, 1)$ and $(0, 1)$, shown in Figure 24. This set can be written as

$$\{(x, y) \in \mathbb{R}^2 : 0 < x < 1,\ 0 < y < 1\}.$$

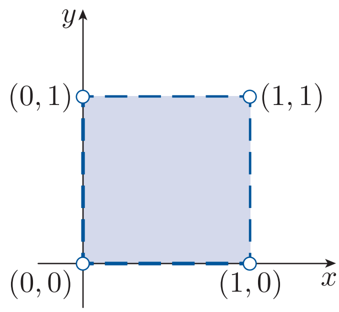


Figure 24 The points inside a square

The square boundary is excluded from this set, and we indicate this by drawing the boundary lines as broken lines and the vertices as *hollow* dots, as in Figure 24.

If we wish our set to include the square boundary, we replace each symbol $<$ by $\leq$, and we indicate this in a sketch by drawing the boundary lines as solid lines and the four vertices as *solid* dots.

These conventions for drawing plane sets are consistent with those you met earlier for intervals. They are summarised below.

Convention for drawing sets in $\mathbb{R}$ or $\mathbb{R}^2$

In a diagram of a subset of $\mathbb{R}$ or $\mathbb{R}^2$:

- included and excluded points are drawn as solid and hollow dots, respectively
- included and excluded boundaries are drawn as solid and broken lines, respectively.

Exercise A16

Sketch each of the following plane sets.

(a) $\{(x, y) \in \mathbb{R}^2 : x < 1\}$

(b) $\{(x, y) \in \mathbb{R}^2 : y \leq 2 - 2x\}$

(c) $\{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + (y - 2)^2 < 4\}$

(d) $\{(x, y) \in \mathbb{R}^2 : x^2 + (y + 3)^2 > 1\}$

Exercise A17

Use set notation to specify the set of points inside the square with vertices $(0, 1)$, $(2, 1)$, $(2, 3)$, $(0, 3)$, together with the boundary, and sketch this set.

2.5 Set equality and subsets

Consider the sets $A = \{1, -1\}$ and $B = \{x \in \mathbb{R} : x^2 - 1 = 0\}$. Although these sets are written in different ways, each set contains exactly the same elements, 1 and -1. We say that these sets are *equal*.

Definition

Two sets A and B are **equal** if they have exactly the same elements; we write $A = B$.

When two sets each contain a small number of elements, we can usually check whether these elements are the same, and hence decide whether the sets are equal.

Exercise A18

Decide whether each of the following is a pair of equal sets.

(a) $A = \{2, -3\}$ and $B = \{x \in \mathbb{R} : x^2 + x - 6 = 0\}$.

(b) $A = \{k \in \mathbb{Z} : k \text{ is odd and } 0 < k < 8\}$ and
$B = \{2n + 1 : n \in \mathbb{N} \text{ and } n^2 < 25\}$.

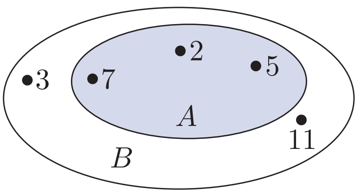


Figure 25 A subset A of a set B

If two sets each contain more than a small number of elements, it is less easy to check whether they are equal. You will meet a method for dealing with cases like this shortly, but first we need the following idea.

Consider the sets $A = \{7, 2, 5\}$ and $B = \{2, 3, 5, 7, 11\}$. These sets are illustrated in the Venn diagram in Figure 25. Each element of A is also an element of B. We say that A is a *subset* of B.

Definition

A set A is a **subset** of a set B if each element of A is also an element of B. We also say that A is *contained in* B, and we write $A \subseteq B$.

Do not confuse the symbol $\subseteq$ with the symbol $\in$. For example, we write

$$\{1\} \subseteq \{1, 2, 3\} \quad \text{and} \quad 1 \in \{1, 2, 3\},$$

because $\{1\}$ is a *subset* of $\{1,2,3\}$ and 1 is an *element* of $\{1,2,3\}$.

We sometimes indicate that a set A is a subset of a set B by reversing the symbol $\subseteq$ and writing $B \supseteq A$, which we read as 'B contains A'.

To indicate that A is *not* a subset of B, we write $A \nsubseteq B$. We may also write this as $B \nsupseteq A$, which we read as 'B does not contain A'.

The next box gives two simple but important facts about subsets.

Subsets of every set

For every set B:

- B is a subset of itself, that is $B \subseteq B$
- the empty set $\varnothing$ is a subset of B, that is, $\varnothing \subseteq B$.

The first result in the box follows immediately from the definition of a subset, given earlier. The second result in the box also follows from the definition, since any set B contains every element of the empty set, for the simple reason that the empty set has no elements.

When we wish to determine whether a set A is a subset of a set B, the method we use depends on the way in which the two sets are defined. If A has a small number of elements, then we can check individually whether each element of A is an element of B. Otherwise, we determine algebraically whether an arbitrary element of A fulfils the membership criteria for B, as illustrated in Worked Exercise A4 below.

To show that a set A is *not* a subset of a set B, we need to find at least one element of A that does not belong to B.

Worked Exercise A4

In each of the following cases, determine whether $A \subseteq B$.

(a) $A = \{1, 2, -4\}$ and $B = \{x \in \mathbb{R} : x^5 + 4x^4 - x - 4 = 0\}$.

(b) $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ and $B = \{(x, y) \in \mathbb{R}^2 : x < 1\}$.

Solution

(a) A has only a few elements, so we can check individually whether they satisfy the membership criteria for B.

The elements 1, 2, -4 belong to $\mathbb{R}$, and

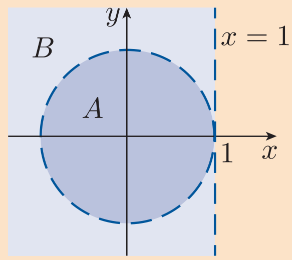
$$1^5 + 4 \times 1^4 - 1 - 4 = 0, \quad \text{so } 1 \in B,$$

$$2^5 + 4 \times 2^4 - 2 - 4 = 90, \quad \text{so } 2 \notin B.$$

Hence A is not a subset of B, that is, $A \nsubseteq B$.

(b) For plane sets, a sketch is often helpful.

The sets A and B are sketched below.



It appears that $A \subseteq B$ but we cannot check each element of A individually, so we try to prove this algebraically.

Let (x, y) be an arbitrary element of A; then (x, y) is a point in $\mathbb{R}^2$ with $x^2 + y^2 < 1$.

Since $y^2 \geq 0$, this implies that $x^2 < 1$, and hence that $x < 1$. Thus $(x, y) \in B$.

Since (x, y) is an arbitrary element of A, we conclude that A is a subset of B, that is, $A \subseteq B$.

Exercise A19

In each of the following cases, determine whether $A \subseteq B$.

(a) $A = \{(5, 2), (1, 1), (-3, 0)\}$ and $B = \{(x, y) \in \mathbb{R}^2 : x - 4y = -3\}$.

(b) $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ and $B = \{(x, y) \in \mathbb{R}^2 : y < 0\}$.

(c) $A = [-1, 0]$ and $B = \{x \in \mathbb{R} : (x + 1)^2 \leq 1\}$.

If a set A is a subset of a set B that is not equal to B, then we say that A is a **proper subset** of B, and we write $A \subset B$ or $B \supset A$.

In some texts, the symbol $\subset$ is used to mean 'is a subset of' (for which we use the symbol $\subseteq$) rather than 'is a proper subset of'.

To show that a set A is a proper subset of a set B, we must show both that A is a subset of B, and that there is at least one element of B that is not an element of A.

Worked Exercise A5

Show that A is a proper subset of B, where:

$A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ and $B = \{(x, y) \in \mathbb{R}^2 : x < 1\}$.

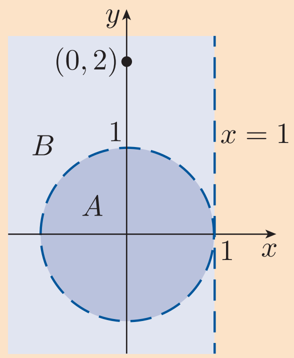
(A and B are the sets you met in Worked Exercise A4(b).)

Solution

We showed in the solution to Worked Exercise A4(b) that $A \subseteq B$.

A sketch can help us find a point in B but not in A. We must then confirm this algebraically.

The sets A and B are sketched below.



The point $(0, 2)$, for example, lies in B, since its x-coordinate 0 is less than 1, but $(0, 2)$ does not lie in A, since $0^2 + 2^2 = 4 \geq 1$. This shows that A is a proper subset of B; that is, $A \subset B$.

Exercise A20

In each of the following cases show that A is a proper subset of B.

(a) $A = \{(5, 2), (1, 1), (-3, 0)\}$ and $B = \{(x, y) \in \mathbb{R}^2 : x - 4y = -3\}$.

(b) $A = [-1, 0]$ and $B = \{x \in \mathbb{R} : (x + 1)^2 \leq 1\}$.

(These sets are the same as those in Exercise A19(a) and (c).)

We now return to the question of how we can show that two sets A and B are equal if they have more than a small number of elements.

If A is a subset of B, we have seen that A is either a *proper* subset of B or is equal to B. Similarly, if B is a subset of A, then B is either a proper subset of A or is equal to A. It follows that, if A is a subset of B *and* B is a subset of A, then the two sets A and B must be equal. This gives us our strategy.

Strategy A1

To show that the sets A and B are equal:

- first show that $A \subseteq B$
- then show that $B \subseteq A$.

Worked Exercise A6

Show that the following sets are equal:

$$A = \{(\cos t, \sin t) : t \in [0, 2\pi]\} \quad \text{and}$$
$$B = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

Solution

We could specify A by
$A = \{(x, y) \in \mathbb{R}^2 : x = \cos t,\ y = \sin t \text{ for some } t \in [0, 2\pi]\}$.

First we show that $A \subseteq B$.

Let (x, y) be an arbitrary element of A; then (x, y) is a point in $\mathbb{R}^2$.

We have $x = \cos t$ and $y = \sin t$, for some $t \in [0, 2\pi]$. So

$$x^2 + y^2 = \cos^2 t + \sin^2 t = 1.$$

This implies that $(x, y) \in B$, so $A \subseteq B$.

Next we show that $B \subseteq A$.

Let (x, y) be an arbitrary element of B; then

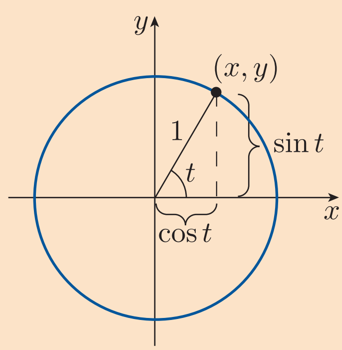
$$x^2 + y^2 = 1.$$

So (x, y) lies on the unit circle.

To show that (x, y) is an element of A, we need to find an angle $t \in [0, 2\pi]$ such that $(x, y) = (\cos t, \sin t)$. A sketch will help.

If we take t to be the (anticlockwise) angle from the (positive) x-axis to the line joining the point (x, y) with the origin, then $t \in [0, 2\pi]$, and

$$x = \cos t \quad \text{and} \quad y = \sin t.$$



It follows that $(x, y) \in A$, so $B \subseteq A$.

Since $A \subseteq B$ and $B \subseteq A$, it follows that $A = B$.

Exercise A21

In each of the following cases, show that the sets A and B are equal.

(a) $A = \{(t^2, 2t) : t \in \mathbb{R}\}$ and $B = \{(x, y) \in \mathbb{R}^2 : y^2 = 4x\}$.

(b) $A = \{(x, y) \in \mathbb{R}^2 : 2x + y - 3 = 0\}$ and $B = \{(t + 1, 1 - 2t) : t \in \mathbb{R}\}$.

2.6 Set operations

Consider the two sets $\{2, 3, 5\}$ and $\{1, 2, 5, 8\}$. Using these sets, we can construct several new sets – for example:

- the set $\{1, 2, 3, 5, 8\}$ consisting of all elements belonging to *at least one* of the two sets
- the set $\{2, 5\}$ consisting of all elements belonging to *both* of the two sets
- the set $\{3\}$ consisting of all elements belonging to the first set but not the second, and the set $\{1, 8\}$ consisting of all elements belonging to the second set but not the first.

Each of these new sets is a particular instance of a general construction for sets. We now consider them in turn.

Union

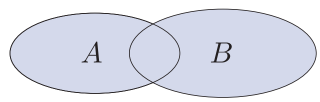
You saw above that if $A = \{2, 3, 5\}$ and $B = \{1, 2, 5, 8\}$, then the set of all elements belonging to at least one of the sets A and B is $\{1, 2, 3, 5, 8\}$. We call this set the *union* of A and B.

More generally, we have the following definition, which is illustrated by the Venn diagram in Figure 26.

Definition

Let A and B be any two sets; then the **union** of A and B is the set

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$



$A \cup B$

Figure 26 The union of sets A and B

The word *or* in this definition is used in the *inclusive* sense of 'and/or'; that is, the set $A \cup B$ consists of the elements of A and the elements of B, including the elements in both A and B. In everyday language, an example of 'or' used in the *exclusive* sense is 'Tea or coffee?', since the answer 'Both, please!' is not expected. An example of 'or' used in the *inclusive* sense is 'Milk or sugar?', since in this case you could answer 'Both'.

Worked Exercise A7

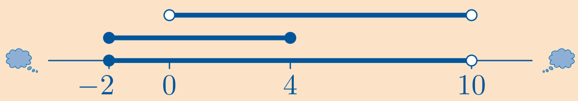
(a) Simplify $[-2, 4] \cup (0, 10)$.

(b) Sketch a diagram depicting the union of the half-plane H and the disc D, where

$$H = \{(x, y) \in \mathbb{R}^2 : y \leq 2 - 2x\},$$
$$D = \{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + (y - 2)^2 < 4\}.$$

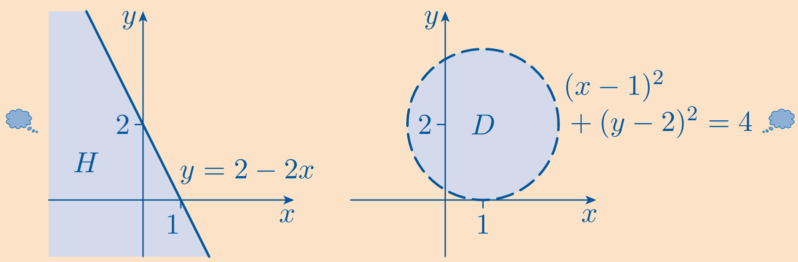
Solution

(a) These intervals overlap.



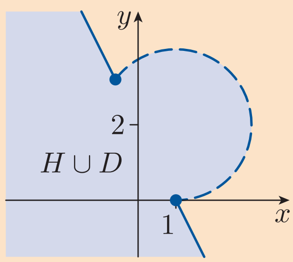
We have $[-2, 4] \cup (0, 10) = [-2, 10)$.

(b) These are the half-plane and disc from Exercise A16(b) and (c).



The union consists of all the points in H or D or both; the two points where the circle and line meet are both in the set H and so are both in the union $H \cup D$ and are shown as solid dots.

The set $H \cup D$ is as follows.



When sketching a set such as that in Worked Exercise A7(b), you should include enough detail so that the set is clear, and therefore the axes and an indication of scale are essential. Finding the exact points where the circle and line meet is not required, but can sometimes be helpful. In this case, substituting $y = 2 - 2x$ into the equation for the circle gives

$$(x-1)^2 + (-2x)^2 = 4,$$

which simplifies to

$$5x^2 - 2x - 3 = 0.$$

This factorises as

$$(x-1)(5x+3) = 0,$$

which has solutions $x = 1$ and $x = -\frac{3}{5}$, so the circle and line meet at the two points $(1, 0)$ and $\left(-\frac{3}{5}, 3\frac{1}{5}\right)$.

Exercise A22

(a) Simplify $(1, 7) \cup [4, 11]$.

(b) Express the set $\mathbb{R}^*$ as a union of intervals.

(c) Sketch a diagram depicting the union of the half-plane H and disc D, where

$$H = \{(x, y) \in \mathbb{R}^2 : y < 0\},$$
$$D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}.$$

So far you have seen the definition of the union of two sets. There is a similar definition for the union of any number of sets; for example, the union of three sets A, B and C, as illustrated by the Venn diagram in Figure 27, is the set

$$A \cup B \cup C = \{x : x \in A \text{ or } x \in B \text{ or } x \in C\}.$$

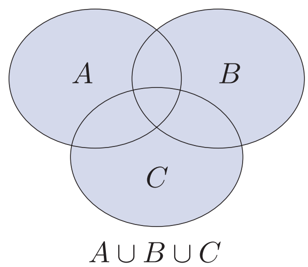


Figure 27 The union of sets A, B and C

Intersection

You saw above that if $A = \{2, 3, 5\}$ and $B = \{1, 2, 5, 8\}$, then the set of all elements belonging to both set A and set B is $\{2, 5\}$. We call this set the *intersection* of A and B.

More generally, we have the following definition, which is illustrated by the Venn diagram in Figure 28.

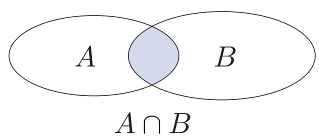


Figure 28 The intersection of sets A and B

Definition

Let A and B be any two sets; then the **intersection** of A and B is the set

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Two sets with no element in common, such as $\{1, 3, 5\}$ and $\{2, 9\}$, are said to be **disjoint**. We write this as $\{1, 3, 5\} \cap \{2, 9\} = \varnothing$ since this intersection is empty.

Worked Exercise A8

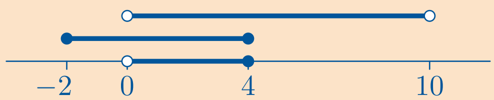
(a) Simplify $[-2, 4] \cap (0, 10)$.

(b) Sketch a diagram depicting the intersection of the half-plane H and disc D, where

$$H = \{(x, y) \in \mathbb{R}^2 : y \leq 2 - 2x\},$$
$$D = \{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + (y - 2)^2 < 4\}.$$

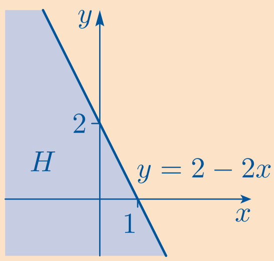
Solution

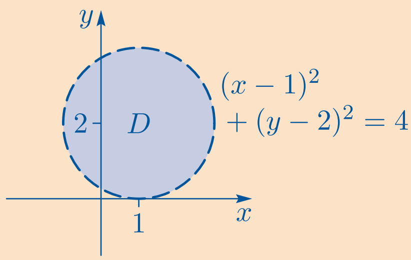
(a) The intersection is the overlap of these intervals.



We have $[-2, 4] \cap (0, 10) = (0, 4]$.

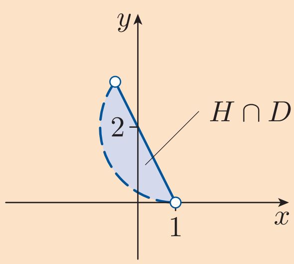
(b) These are the half-plane and disc from Exercise A16(b) and (c), and Worked Exercise A7.





The intersection consists of all the points in both H and D. Neither of the points where the circle and the line meet are in the set D, so these points are not in the intersection $H \cap D$, and both are shown as hollow dots.

The set $H \cap D$ is as follows.



Exercise A23

(a) Simplify $(1,7) \cap [4,11]$.

(b) Sketch a diagram depicting the intersection of the half-plane H and disc D, where

$$H = \{(x,y) \in \mathbb{R}^2 : y < 0\},$$
$$D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}.$$

(These are the same sets as in Exercise A22(a) and (c).)

So far you have seen the definition of the intersection of two sets. There is a similar definition for the intersection of any number of sets; for example, the intersection of three sets A, B and C, as illustrated by the Venn diagram in Figure 29, is the set

$$A \cap B \cap C = \{x : x \in A \text{ and } x \in B \text{ and } x \in C\}.$$

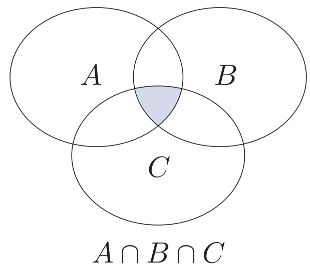


Figure 29 The intersection of sets A, B and C

Difference

You saw above that if $A = \{2,3,5\}$ and $B = \{1,2,5,8\}$, then the set of all elements belonging to A but not to B is $\{3\}$; we call this set the *difference* $A - B$. Similarly, the set of all elements belonging to B but not to A is $\{1,8\}$; this set is the *difference* $B - A$.

More generally, we have the following definition, which is illustrated by the Venn diagram in Figure 30.

Definition

Let A and B be any two sets; then the **difference** between A and B is the set

$$A - B = \{x : x \in A,\ x \notin B\}.$$

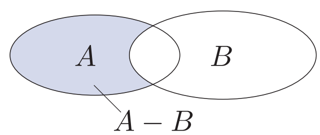


Figure 30 The difference between set A and set B

Notice that $A - B$ is different from $B - A$ when $A \neq B$. This is unlike the union and intersection, where $A \cup B = B \cup A$ and $A \cap B = B \cap A$, for any sets A and B. Also, for any set A, we have $A - A = \varnothing$, again unlike the union and intersection, where $A \cup A = A \cap A = A$.

In some texts the difference $A - B$ of two sets A and B is denoted by $A \setminus B$.

Worked Exercise A9

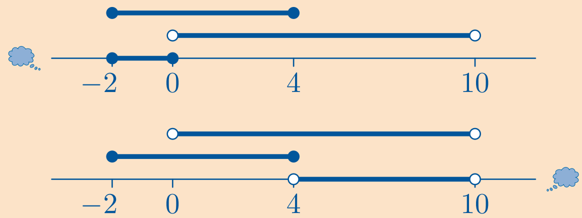
(a) Simplify $[-2, 4] - (0, 10)$ and $(0, 10) - [-2, 4]$.

(b) Sketch diagrams depicting the differences $H - D$ and $D - H$ of the half-plane H and disc D, where

$$H = \{(x, y) \in \mathbb{R}^2 : y \leq 2 - 2x\},$$
$$D = \{(x, y) \in \mathbb{R}^2 : (x-1)^2 + (y-2)^2 < 4\}.$$

Solution

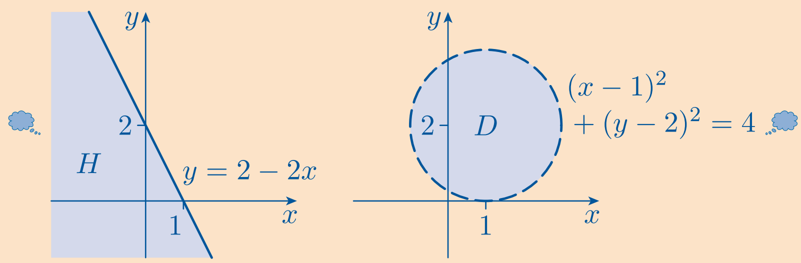
(a)



We have

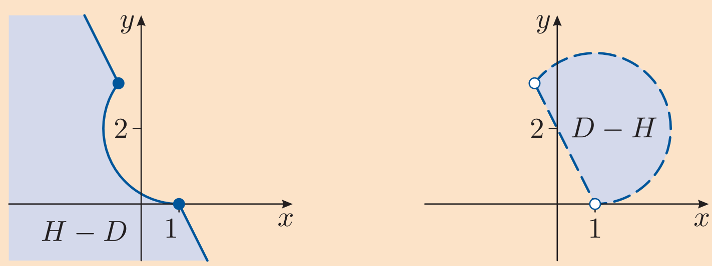
$$[-2, 4] - (0, 10) = [-2, 0],$$
$$(0, 10) - [-2, 4] = (4, 10).$$

(b) Again these are the half-plane and disc from Exercise A16(b) and (c), and Worked Exercises A7 and A8.



Consider carefully the boundary points, and in particular, the points where the line and circle meet. Both of the meeting points are in $H - D$, as are the remaining points of the boundaries. Neither of the meeting points is in the difference $D - H$, nor are the remaining points of the boundaries.

The sets $H - D$ and $D - H$ are as follows.



Exercise A24

(a) Simplify $(1,7) - [4,11]$ and $[4,11] - (1,7)$.

(b) Sketch diagrams depicting the differences $H - D$ and $D - H$ of the half-plane H and disc D, where

$$H = \{(x,y) \in \mathbb{R}^2 : y < 0\},$$
$$D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}.$$

(These are the same sets as in Exercise A22(a) and (c), and Exercise A23.)

Gottfried Wilhelm Leibniz

3 Functions

In this section you will revise what is meant by a *function*, and some associated ideas. You will look at not only functions of real numbers, but also functions of other mathematical objects. The idea of a function is fundamental throughout this module, so it is vital that you have a good understanding of this topic.

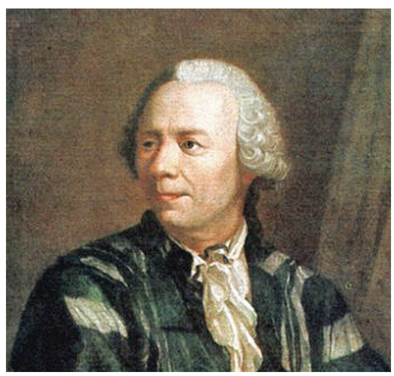
The term ‘function’ first emerged at the end of the seventeenth century in the correspondence of Gottfried Wilhelm Leibniz (1646–1716) and Johann Bernoulli (1667–1748). But it was Leonhard Euler (1707–1783) in the middle of the eighteenth century who was responsible for the essential development, notably through his *Introductio in Analysin Infinitorum* of 1748, the first work in which the concept of a function plays an explicit and central role.

Johann Bernoulli

3.1 What is a function?

You can think of a *function* as a machine for processing mathematical objects, such as numbers, points in the plane or vectors.

For example, consider the function f that takes non-zero real numbers as its inputs and whose rule is that the input x leads to the output $f(x) = 1/x$. You can regard it as a machine that calculates the reciprocals of its input numbers. When 3 is fed into the machine, out comes $\frac{1}{3}$; when -2 is fed into the machine, out comes $-\frac{1}{2}$; and so on. Any real number in the *domain* $\mathbb{R}^*$ of f can be processed by the machine to produce a real number in the *codomain* $\mathbb{R}$ of f, as illustrated in Figure 31.



Leonhard Euler

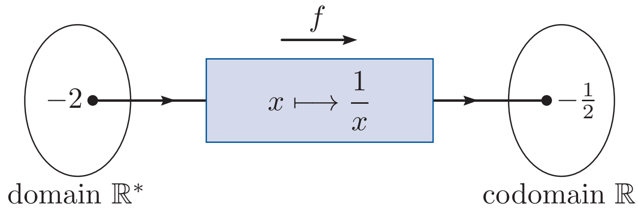


Figure 31 A function as a machine

Similarly, consider the function g that accepts points in the plane as its inputs and whose rule is that the input (x, y) leads to the output $g((x, y)) = y$. You can regard it as a machine that calculates the y-coordinate of each input point. When the point $(1, 2)$ is fed into the machine, out comes 2; when the point $(0, 0)$ is fed into the machine, out comes 0; and so on. Any point in the domain $\mathbb{R}^2$ of g can be processed by the machine to produce a real number in the codomain $\mathbb{R}$ of g, as illustrated in Figure 32.

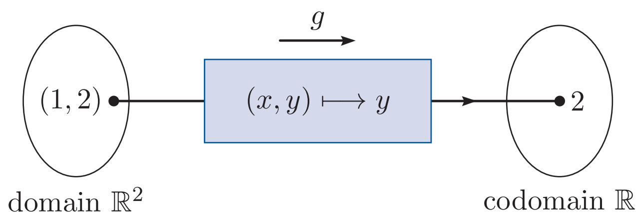


Figure 32 Another function as a machine

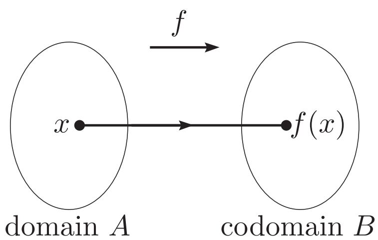


Figure 33 A general function

In general, imagine a machine that accepts an element x from some set A, and processes it to produce a single element $f(x)$ in some set B. This machine corresponds to the following general definition of a function, which is illustrated in Figure 33.

Definition

A **function** f is defined by specifying:

- a set A, called the **domain** of f
- a set B, called the **codomain** of f
- a **rule** $x \longmapsto f(x)$ that associates each element $x \in A$ with a unique element $f(x) \in B$.

The element $f(x)$ is the **image** of x under f.

Symbolically, we write

$$\begin{aligned} f : A &\longrightarrow B \\ x &\longmapsto f(x). \end{aligned}$$

We often refer to a function as a **mapping**, and say that f **maps** A to B and x to $f(x)$.

Notice that the definition of a function does not require *every* element of the codomain B to be the image of an element of the domain A, but it *does* require every element of the domain A to have an image in the codomain B. For example, a function with rule $x \longmapsto \sin x$ and domain $\mathbb{R}$ could have codomain $\mathbb{R}$, or $[-1, 1]$, or any set of real numbers of which $[-1, 1]$ is a subset, but not, say, codomain $[0, 1]$ since the image of $3\pi/2$ is $\sin(3\pi/2) = -1$, which is not in this set.

Notice also that the symbolic definition of a function given at the end of the box above specifies all three of the constituent parts of a function at once: the domain, the codomain and the rule. For example, the definition

$$\begin{aligned} f : \mathbb{Z} &\longrightarrow \mathbb{Z} \\ n &\longmapsto n + 1 \end{aligned}$$

specifies a function with domain $\mathbb{Z}$, codomain $\mathbb{Z}$ and rule $f(n) = n + 1$.

When we write a function symbolically, the first arrow is unbarred to signify a mapping from the domain A to the codomain B. The second arrow is barred, to show that the *particular* element x of A is mapped to the *particular* element $f(x)$ of B. Each arrow is read as 'maps to'.

The following paragraphs give a number of examples of different types of functions.

Real functions

A function whose domain and codomain are both subsets of $\mathbb{R}$ is called a **real function**. Examples include the functions

$$\begin{aligned} f : \mathbb{R}^* &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{x} \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 2x - 5. \end{aligned}$$

In some texts, a real function is defined to be a function whose codomain is a subset of $\mathbb{R}$, but whose domain can be any set.

You may be more familiar with seeing these functions written as simply $f(x) = 1/x$ and $g(x) = 2x - 5$. We write functions in this shortened way when it is understood from the context what the domain and codomain are.

Distance function

Functions of the form $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ can be used to specify quantities associated with points in the plane. For example, the function

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \sqrt{x^2 + y^2} \end{aligned}$$

gives the distance of each point (x, y) in the plane from the origin, as shown in Figure 34.

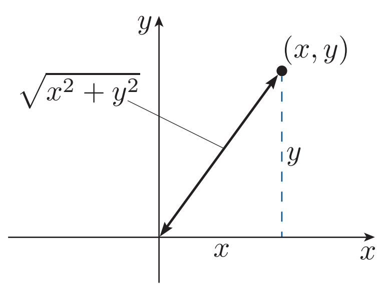


Figure 34 The distance of a point from the origin

Transformations of the plane

Functions that have a geometric interpretation are often called **transformations**. Such functions include translations, reflections and rotations of the plane. We now look at some simple examples. For each one, the diagram shows the effect of the transformation on the square whose vertices are at $(0, 0)$, $(1, 0)$, $(1, 1)$ and $(0, 1)$; part of the square is shaded for clarity.

- The transformation

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x + 2, y) \end{aligned}$$

  is the **translation** of the plane that shifts (or translates) each point to the right by 2 units, as illustrated in Figure 35.

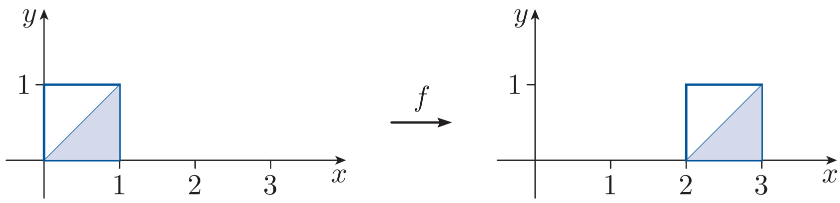


Figure 35 Translation 2 units to the right

- The transformation

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (-x, y) \end{aligned}$$

  is the **reflection** of the plane in the y-axis, as illustrated in Figure 36.

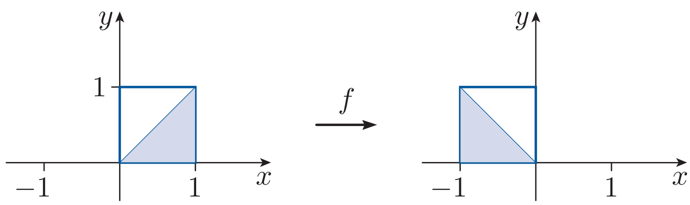


Figure 36 Reflection in the y-axis

- The transformation

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (-x, -y) \end{aligned}$$

  is the **rotation** of the plane through an angle π about the origin, as illustrated in Figure 37.

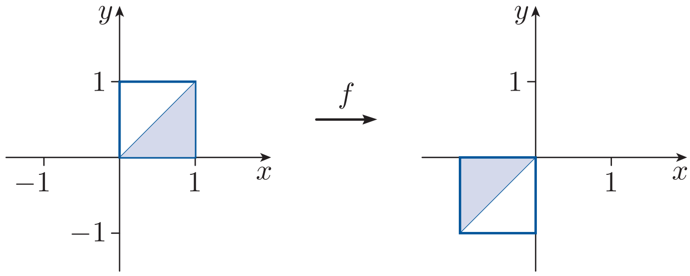


Figure 37 Rotation through an angle π about the origin

When specifying a function, like a transformation, where the elements of the domain are of the form (x, y), we simply write $f(x, y)$ rather than $f((x, y))$.

Exercise A25

For each of the following functions $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, state whether f is a translation, reflection or rotation of the plane.

(a) $f(x, y) = (x + 2, y + 3)$

(b) $f(x, y) = (x, -y)$

(c) $f(x, y) = (-y, x)$

Functions whose domains are finite sets

It is often useful to consider a function whose domain is a *finite* set. For example, we can define a function whose domain and codomain are the set

$$A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

by

$$\begin{aligned} f : A &\longrightarrow A \\ x &\longmapsto 9 - x. \end{aligned}$$

When the domain of a function f has a small number of elements, we can specify the rule of f by listing the image $f(x)$ of each element x in the domain. For example, let $A = \{0, 1, 2, 3\}$ and $B = \{2, 3, 4, 5\}$; then we can define a function $f : A \longrightarrow B$ by the rule

$$f(0) = 2, \quad f(1) = 2, \quad f(2) = 4, \quad f(3) = 5.$$

We can represent the behaviour of this function by a diagram, as shown in Figure 38. A diagram of this type that represents a function always has the following features:

- there is *exactly one* arrow *from* each element in the domain, since each element in the domain has *exactly one* image in the codomain
- there may be no arrows, one arrow or several arrows going *to* an element in the codomain, since an element in the codomain may not be an image at all, may be an image of exactly one element in the domain, or may be an image of several elements in the domain.

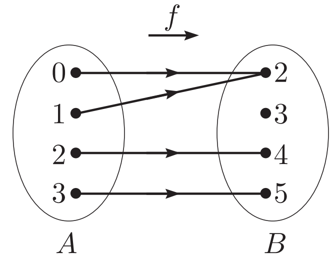
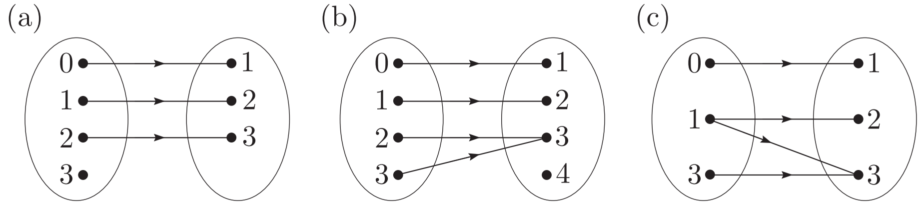


Figure 38 Function f from set A to set B

In the example shown in Figure 38, the number 3 is not an image at all, 5 is the image of 3 only, and 2 is the image of both 0 and 1.

Exercise A26

Which of the following diagrams represent(s) a function?



Identity functions

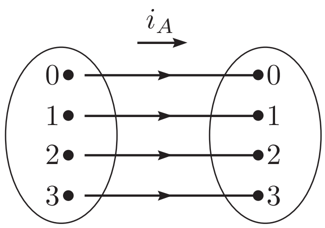


Figure 39 An identity function

Associated with any set A, there is a particularly simple function whose domain and codomain are the set A. This is the identity function i_A, which maps each element of A to itself. (We sometimes omit the subscript A if we do not need to emphasise the set.)

For example, let $A = \{0, 1, 2, 3\}$; then the rule of the identity function i_A, as illustrated in Figure 39, is

$$i_A(0) = 0, \quad i_A(1) = 1, \quad i_A(2) = 2, \quad i_A(3) = 3.$$

The following definition applies to *any* set A, finite or infinite.

Definition

The **identity function** on a set A is the function

$$\begin{aligned} i_A : A &\longrightarrow A \\ x &\longmapsto x. \end{aligned}$$

3.2 Image set of a function

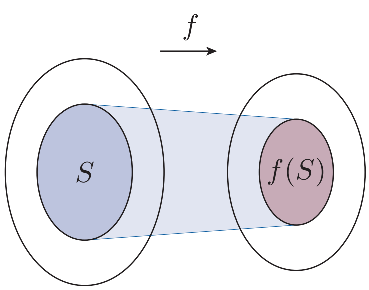


Figure 40 Image of a set S under a function f

The rule associated with a function tells us how to find the image of any element in the domain. Often, however, we need to consider the images of all elements in some subset of the domain. The subset of the codomain containing these images is called the *image* of the original subset, as stated below and illustrated in Figure 40.

Definition

Let $f : A \longrightarrow B$ be a function. For any subset S of A, the **image** of S under f, denoted by $f(S)$, is the set

$$f(S) = \{f(x) : x \in S\}.$$

Worked Exercise A10

Find $f(S)$, where $S = \{1, 2, 3\}$ and

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{x}. \end{aligned}$$

Solution

$f(S) = \{f(1), f(2), f(3)\} = \{1, \frac{1}{2}, \frac{1}{3}\}$.

Exercise A27

Let

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x - 1. \end{aligned}$$

Find the image under f of each of the following sets.

(a) $S = \{0, 1, 2, 3\}$ (b) $\mathbb{Z}$

The idea of the image of a subset of elements is useful in geometry, for example, where we frequently want to consider the effect of a transformation on a plane figure, a subset of $\mathbb{R}^2$. For example, suppose that S is the square with vertices at $(0, 0)$, $(1, 0)$, $(1, 1)$ and $(0, 1)$, and we want to find the image of S under the function

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x + 2, y). \end{aligned}$$

This function is the translation of the plane that moves each point (x, y) to the right by 2. The image of S is therefore the square with vertices at $f(0, 0) = (2, 0)$, $f(1, 0) = (3, 0)$, $f(1, 1) = (3, 1)$ and $f(0, 1) = (2, 1)$, as shown in Figure 41.

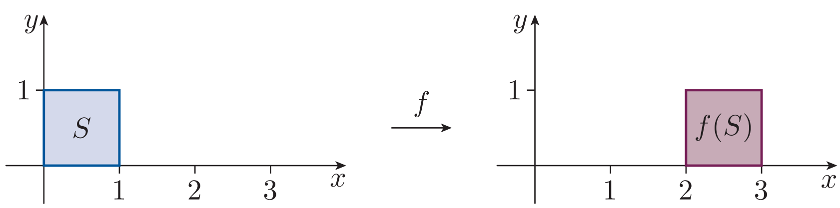


Figure 41 The image $f(S)$ of a square S under a translation f

Sometimes we want to consider the image of the *whole domain* of a function: this set is called the *image set* of the function, as illustrated in Figure 42.

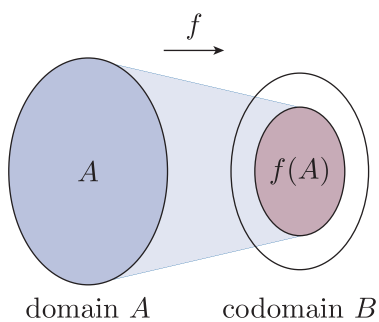


Figure 42 Image set of a function f

Definition

The **image set** of a function $f : A \longrightarrow B$ is the set

$$f(A) = \{f(x) : x \in A\}.$$

The image set of a function is a subset of its codomain. It need not be *equal* to the codomain because there may be some elements of the codomain that are not images of elements in the domain.

In some texts, the image set of a function is called the *image* of the function, or the *range* of the function.

When the domain of a function f has a small number of elements, we can find the image set of f by finding the image of each element in the domain, and listing them to form a set.

Worked Exercise A11

Let $A = \{-3, -2, -1, 0, 1, 2, 3\}$ and $B = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Find the image set of the function

$$\begin{aligned} f : A &\longrightarrow B \\ x &\longmapsto x^2. \end{aligned}$$

Solution

The images of the elements of A are

$$f(-3) = 9, \quad f(-2) = 4, \quad f(-1) = 1, \quad f(0) = 0,$$
$$f(1) = 1, \quad f(2) = 4, \quad f(3) = 9.$$

So the image set of f is $f(A) = \{0, 1, 4, 9\}$.

Exercise A28

Let $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Find the image set of the function

$$\begin{aligned} f : A &\longrightarrow A \\ x &\longmapsto 9 - x. \end{aligned}$$

You should have found that for the particular function in Exercise A28 the image set and the codomain are the same set. In other words, each element of the codomain is the image of an element in the domain, as illustrated in Figure 43. A function with this property is said to be *onto*.

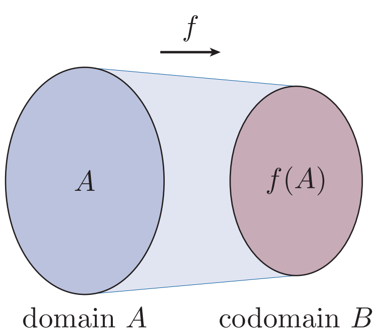


Figure 43 An onto function: $f(A) = B$

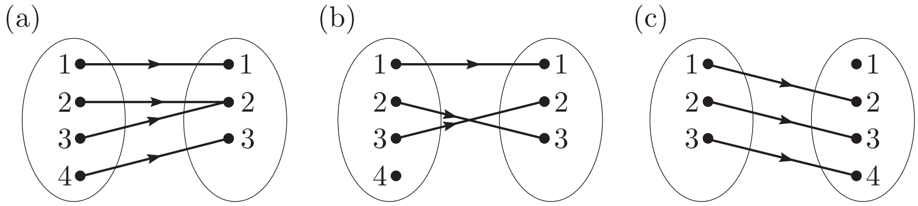
Definition

A function $f : A \longrightarrow B$ is **onto** if $f(A) = B$.

Some texts refer to an onto function as a *surjective* function.

Exercise A29

Which of the following diagrams represent(s) an onto function?



You have seen that if the domain of a function is a small finite set, then we can find the image set of the function by finding the image of each element of the domain individually. If the domain is a large finite set or an infinite set, then we need an algebraic argument to determine the image set. Sometimes we 'guess' what the image set seems to be, and then confirm this algebraically.

For a real function, a sketch of its graph can help us 'guess' the image set. For a function that is a transformation of the plane, we can use our knowledge of such transformations to help us 'guess' the image set.

To show that the image set is equal to our 'guess' set, we use our usual strategy for showing that two sets are equal: we show that each is a subset of the other.

- To show that the image set is a subset of our 'guess' set, we show that the image of an arbitrary element of the domain lies in our 'guess' set.
- To show that our 'guess' set is a subset of the image set, we take an arbitrary element of our 'guess' set and find an element of the domain whose image is this arbitrary element.

Worked Exercise A12

For each of the following functions, find its image set and determine whether it is onto.

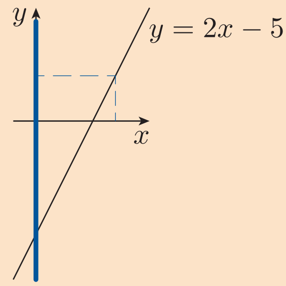
(a) $f : \mathbb{R} \longrightarrow \mathbb{R}$
$\quad x \longmapsto 2x - 5$

(b) $f : \mathbb{R} \longrightarrow \mathbb{R}$
$\quad x \longmapsto x^2$

(c) $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$
$\quad (x, y) \longmapsto (x + 1, y + 2)$

Solution

(a) A sketch of the graph of f is shown below.



For every element on the y-axis, a horizontal line drawn through that element meets the graph. So it seems that every element of the codomain is the image of some element of the domain. That is, we 'guess' that the image set $f(\mathbb{R})$ is the whole codomain $\mathbb{R}$.

We prove that $f(\mathbb{R}) = \mathbb{R}$.

The image set is always a subset of the codomain; in this case the codomain is $\mathbb{R}$, so $f(\mathbb{R}) \subseteq \mathbb{R}$.

We know that $f(\mathbb{R}) \subseteq \mathbb{R}$, so we must show that $f(\mathbb{R}) \supseteq \mathbb{R}$.

We take an arbitrary element in our 'guess' set $\mathbb{R}$, and find an element in the domain $\mathbb{R}$ whose image is this arbitrary element.

Let y be an arbitrary element in $\mathbb{R}$. We must show that $y \in f(\mathbb{R})$; that is, there exists an element x in the domain $\mathbb{R}$ such that

$$f(x) = y; \quad \text{that is,} \quad 2x - 5 = y.$$

Rearranging this equation, we obtain

$$x = \frac{y+5}{2}$$

which is in the domain $\mathbb{R}$. So we have

$$\begin{aligned} f(x) &= 2x - 5 \\ &= 2\left(\frac{y+5}{2}\right) - 5 \\ &= y, \end{aligned}$$

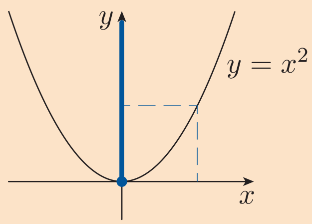
that is, for every $y \in \mathbb{R}$ there is an x in the domain $\mathbb{R}$ such that $f(x) = y$.

Thus $f(\mathbb{R}) \supseteq \mathbb{R}$.

Since $f(\mathbb{R}) \subseteq \mathbb{R}$ and $f(\mathbb{R}) \supseteq \mathbb{R}$, it follows that $f(\mathbb{R}) = \mathbb{R}$, so the image set of f is indeed $\mathbb{R}$.

The codomain of f is also $\mathbb{R}$, so f is onto.

(b) A sketch of the graph of f is shown below.



For every element in the interval $[0, \infty)$ of the y-axis (marked on the sketch), a horizontal line drawn through that element meets the graph. For any element outside this interval, such a horizontal line does not meet the graph. So we 'guess' that the image set $f(\mathbb{R})$ is $[0, \infty)$.

We prove that $f(\mathbb{R}) = [0, \infty)$.

We know that the image set is a subset of the codomain $\mathbb{R}$, but we don't know that it is a subset of $[0, \infty)$. We have to show algebraically that $f(\mathbb{R}) \subseteq [0, \infty)$ by finding the image of an arbitrary element in the domain $\mathbb{R}$.

Let x be an arbitrary element in the domain $\mathbb{R}$; then $f(x) = x^2$. Now, $x^2 \geq 0$ for all $x \in \mathbb{R}$, so $f(\mathbb{R}) \subseteq [0, \infty)$.

We must now show that $f(\mathbb{R}) \supseteq [0, \infty)$.

We take an arbitrary element in our 'guess' set $[0, \infty)$, and find an element of the domain $\mathbb{R}$ whose image is this arbitrary element.

Let y be an arbitrary element in $[0, \infty)$. We must show that there exists an element x in the domain $\mathbb{R}$ such that

$$f(x) = y; \quad \text{that is,} \quad x^2 = y.$$

Now $x = \sqrt{y}$ is in $\mathbb{R}$ (since $y \geq 0$) and satisfies $f(x) = y$, as required. Thus $f(\mathbb{R}) \supseteq [0, \infty)$.

Since $f(\mathbb{R}) \subseteq [0, \infty)$ and $f(\mathbb{R}) \supseteq [0, \infty)$, it follows that $f(\mathbb{R}) = [0, \infty)$, so the image set of f is $[0, \infty)$.

The image set $f(\mathbb{R}) = [0, \infty)$ is not the whole of the codomain $\mathbb{R}$, so f is not onto.

If we had simply been asked to determine whether f is onto, we could have shown that it is not by finding just one element, say -1, in the codomain $\mathbb{R}$ that is not the image of an element of the domain $\mathbb{R}$.

(c) This function is a translation of the plane (it shifts each point to the right by 1 unit and up by 2 units). So we expect ('guess') the image set to be the plane $\mathbb{R}^2$.

We prove that $f(\mathbb{R}^2) = \mathbb{R}^2$.

The image set is always a subset of the codomain; in this case the codomain is $\mathbb{R}^2$, so $f(\mathbb{R}^2) \subseteq \mathbb{R}^2$.

We know that $f(\mathbb{R}^2) \subseteq \mathbb{R}^2$, so we must show that $f(\mathbb{R}^2) \supseteq \mathbb{R}^2$.

Let (x', y') be an arbitrary element in the codomain $\mathbb{R}^2$. We must show that there exists an element (x, y) in the domain $\mathbb{R}^2$ such that

$$f(x,y) = (x', y'); \quad \text{so,} \quad (x+1, y+2) = (x', y'),$$

that is, $x + 1 = x'$ and $y + 2 = y'$. Rearranging these two equations, we obtain

$$x = x' - 1, \quad y = y' - 2.$$

So, $(x, y) \in \mathbb{R}^2$ and $f(x, y) = (x', y')$, as required. Thus $f(\mathbb{R}^2) \supseteq \mathbb{R}^2$.

Since $f(\mathbb{R}^2) \subseteq \mathbb{R}^2$ and $f(\mathbb{R}^2) \supseteq \mathbb{R}^2$, it follows that $f(\mathbb{R}^2) = \mathbb{R}^2$, so the image set of f is $\mathbb{R}^2$.

The codomain of f is also $\mathbb{R}^2$, so f is onto.

Exercise A30

For each of the following functions, find its image set and determine whether it is onto.

(a) $\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 1 + x^2 \end{aligned}$ (b) $\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x, -y) \end{aligned}$

As you have seen from Worked Exercise A12 and Exercise A30, when you want to determine whether a function is onto, it is crucial to take into account what the *codomain* of the function is. For example, you saw in Worked Exercise A12 that the function

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned}$$

is not onto. To see this, you just have to observe that the element -1, for example, of the codomain is not the image of any element of the domain. However, if you remove all the negative numbers from the codomain of this function, then you obtain the new function

$$\begin{aligned} g : \mathbb{R} &\longrightarrow [0, \infty) \\ x &\longmapsto x^2, \end{aligned}$$

which is onto, since every element of the codomain is an image of an element of the domain. Note that these functions f and g are *different* functions, since they have different codomains.

3.3 Inverse functions

Given a function

$$\begin{aligned} f : A &\longrightarrow B \\ x &\longmapsto f(x), \end{aligned}$$

it is sometimes possible to define an *inverse function* that 'undoes' the effect of f by mapping each image element $f(x)$ back to the element x whose image it is. For example, a rotation in the plane can be 'undone' by a rotation in the opposite direction.

However, consider the function

$$\begin{aligned} f : A &\longrightarrow B \\ x &\longmapsto x^2, \end{aligned}$$

where $A = \{-3, -2, -1, 0, 1, 2, 3\}$ and $B = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

We know that $f(-2) = f(2) = 4$, and so a function that 'undoes' the effect of f must map the number 4 to the number -2 *and* to the number 2, which is impossible. Thus, in this case, no inverse function exists. This function f is an example of a function that is *many-to-one*. A many-to-one function does not have an inverse function.

Definitions

A function $f : A \longrightarrow B$ is **one-to-one** if each element of $f(A)$ is the image of exactly one element of A; that is,

$$\text{if } x_1, x_2 \in A \text{ and } f(x_1) = f(x_2), \quad \text{then } x_1 = x_2.$$

A function that is not one-to-one is **many-to-one**.

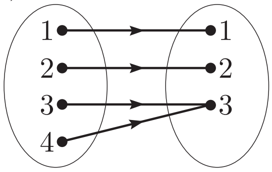
Thus a function f is one-to-one if it maps distinct elements in the domain A to distinct elements in the image set $f(A)$. Some texts refer to a one-to-one function as an *injective* function.

To prove that a function f is *not* one-to-one (that is, that the function is many-to-one), it is sufficient to find just one pair of *distinct* elements in the domain A with the *same* image under f.

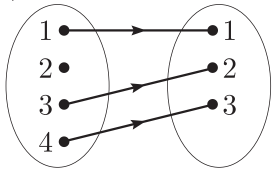
Exercise A31

Which of the following diagrams represent(s) a one-to-one function?

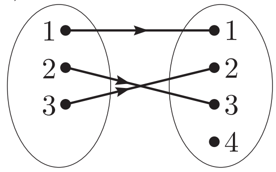
(a)



(b)



(c)



If the domain of a function is a large finite set or an infinite set, then to show that the function is one-to-one, we need an algebraic argument. We aim to show algebraically that, if two elements of the domain have the same image under the function, then they must actually be the same element, as demonstrated in Worked Exercise A13.

Showing that a function is *not* one-to-one is more straightforward: we just give a pair of distinct elements that have the same image under the function, as you have seen.

For a real function, an initial sketch of its graph can help us 'guess' whether or not the function is one-to-one, and if it is not one-to-one, the graph can also help us find a pair of elements that show this.

Worked Exercise A13

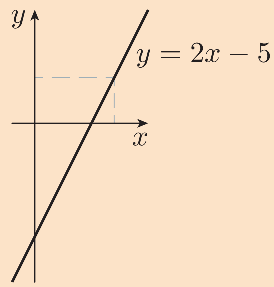
Determine which of the following functions are one-to-one.

(a) $f : \mathbb{R} \longrightarrow \mathbb{R}$, $x \longmapsto 2x - 5$ (b) $f : \mathbb{R} \longrightarrow \mathbb{R}$, $x \longmapsto x^2$ (c) $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, $(x, y) \longmapsto (x + 1, y + 2)$

(These are the same functions as in Worked Exercise A12.)

Solution

(a) A sketch of the graph of f is shown below.



Each horizontal line meets the graph just once. So it seems that no element of the codomain is the image of more than one element of the domain. That is, it seems that f is one-to-one. To prove this, we show that if two elements x_1 and x_2 in the domain have the same image, then they must actually be the *same* element.

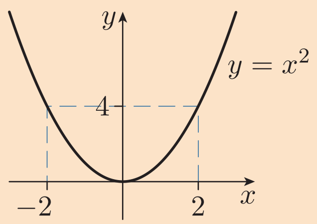
We show that f is one-to-one. Suppose that $f(x_1) = f(x_2)$; then

$$2x_1 - 5 = 2x_2 - 5,$$

so $2x_1 = 2x_2$, and hence $x_1 = x_2$.

Thus f is one-to-one.

(b) A sketch of the graph of f is shown below.



Some horizontal lines meet the graph more than once. So it seems that f is not one-to-one. To show this, we find two distinct elements of the domain with the same image.

This function is not one-to-one since, for example,

$$f(2) = f(-2) = 4.$$

(c) This function is a translation of the plane, so we expect it to be one-to-one.

We show that f is one-to-one. Suppose that $f(x_1, y_1) = f(x_2, y_2)$; then

$$(x_1 + 1, y_1 + 2) = (x_2 + 1, y_2 + 2).$$

Thus

$$x_1 + 1 = x_2 + 1 \quad \text{and} \quad y_1 + 2 = y_2 + 2,$$

so

$$x_1 = x_2 \quad \text{and} \quad y_1 = y_2.$$

Hence $(x_1, y_1) = (x_2, y_2)$, so f is one-to-one.

Exercise A32

Determine which of the following functions is one-to-one.

(a) $f : \mathbb{R} \longrightarrow \mathbb{R}$
$\quad x \longmapsto 1 + x^2$

(b) $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$
$\quad (x, y) \longmapsto (x, -y)$

(These are the same functions as in Exercise A30.)

For a one-to-one function $f : A \longrightarrow B$, we have the situation illustrated in Figure 44. Each element y in $f(A)$ is the image of a unique element x in A, and so we can reverse the arrows to obtain the *inverse function* with domain $f(A)$ and image set A, which maps y back to x. When it exists, we denote the inverse function of f by f^{-1}.

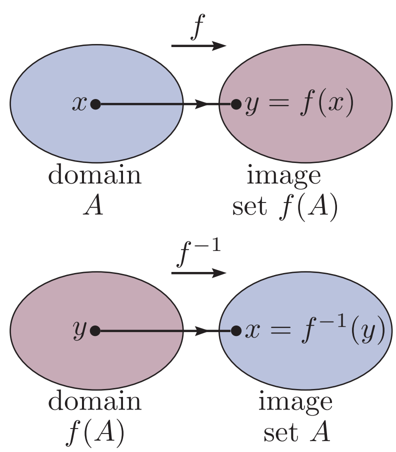


Figure 44 A function and its inverse

Definition

Let $f : A \longrightarrow B$ be a one-to-one function. Then f has an **inverse function** $f^{-1} : f(A) \longrightarrow A$, with rule

$$f^{-1}(y) = x, \quad \text{where } y = f(x).$$

Notice in this definition that the domain of f^{-1} is $f(A)$; it is not necessarily the whole of B.

However, if a function $f : A \longrightarrow B$ is *onto*, as well as one-to-one, then f has an inverse function $f^{-1} : B \longrightarrow A$; that is, the domain of f^{-1} *is* the whole of B.

A function $f : A \longrightarrow B$ that is both one-to-one and onto is said to be a **one-to-one correspondence** between the sets A and B. For such a function f, not only is f^{-1} the inverse of f, but also f is the inverse of f^{-1}; that is, the functions f and f^{-1} are inverses of each other.

Some texts refer to a one-to-one correspondence as a *bijection*.

Worked Exercise A14

For each of the following functions, determine whether f has an inverse function f^{-1}; if it exists, find it.

(a) $\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 2x - 5 \end{aligned}$

(b) $\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned}$

(c) $\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x + 1, y + 2) \end{aligned}$

(d) $\begin{aligned} f : [0, \infty) &\longrightarrow [-1, \infty) \\ x &\longmapsto 3x^2 - 1 \end{aligned}$

Solution

(a) In Worked Exercise A13(a), we showed that f is one-to-one, so f has an inverse function.

In Worked Exercise A12(a), we showed that the image set of f is $\mathbb{R}$ and that, for each y in the image set $\mathbb{R}$, there is an $x \in \mathbb{R}$ such that

$$y = f(x) = f\left(\frac{y + 5}{2}\right).$$

Under f, we know that y is the image of $(y + 5)/2$, so under the inverse, $(y + 5)/2$ is the image of y.

So f^{-1} is the function

$$\begin{aligned} f^{-1} : \mathbb{R} &\longrightarrow \mathbb{R} \\ y &\longmapsto \frac{y + 5}{2}. \end{aligned}$$

It does not matter whether the definition of f^{-1} is expressed in terms of x or y, but it is more usual to use x in the definition of a real function.

This definition can be expressed in terms of x as

$$\begin{aligned} f^{-1} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{x+5}{2}. \end{aligned}$$

(b) In Worked Exercise A13(b), we showed that f is not one-to-one, so f does not have an inverse function.

(c) In Worked Exercise A13(c), we showed that f is one-to-one, so f has an inverse function.

In Worked Exercise A12(c), we showed that the image set of f is $\mathbb{R}^2$ and that, for each (x', y') in the image set $\mathbb{R}^2$, we have

$$(x', y') = f(x, y) = f(x' - 1, y' - 2).$$

Under f, we know that (x', y') is the image of $(x' - 1, y' - 2)$, so under the inverse, $(x' - 1, y' - 2)$ is the image of (x', y').

So f^{-1} is the function

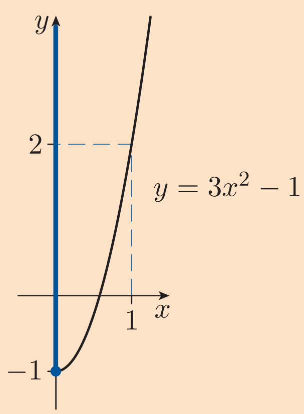
$$\begin{aligned} f^{-1} : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x', y') &\longmapsto (x' - 1, y' - 2). \end{aligned}$$

This definition can be expressed in terms of x and y as

$$\begin{aligned} f^{-1} : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x - 1, y - 2). \end{aligned}$$

This makes sense: geometrically, f is the translation that shifts each point to the right by 1 unit and up by 2 units, so we expect the inverse to be a translation to the left by 1 unit and down by 2 units.

(d) A sketch of the graph of f is shown below.



Each horizontal line meets the graph just once. So it seems that f is one-to-one. To prove this, we show that if two elements x_1 and x_2 in the domain have the same image, then they must actually be the *same* element.

We show that f is one-to-one. Suppose that $f(x_1) = f(x_2)$; then

$$3x_1^2 - 1 = 3x_2^2 - 1,$$

so $3x_1^2 = 3x_2^2$, and hence $x_1^2 = x_2^2$. Since both x_1 and x_2 are in the domain $[0, \infty)$, this implies that $x_1 = x_2$.

Thus f is one-to-one.

We now find the image set of f. From the sketch, we 'guess' that it is $[-1, \infty)$, the codomain of f. That is, we guess that f is onto.

We prove that $f([0, \infty)) = [-1, \infty)$.

The image set is a subset of the codomain.

We know that $f([0, \infty)) \subseteq [-1, \infty)$, so we must show that $f([0, \infty)) \supseteq [-1, \infty)$.

We take an arbitrary element in our 'guess' set $[-1, \infty)$, and find an element of the domain $[0, \infty)$ whose image is this arbitrary element.

Let y be an arbitrary element in $[-1, \infty)$. We must show that there exists an element x in the domain $[0, \infty)$ such that

$$f(x) = y; \quad \text{that is,} \quad 3x^2 - 1 = y.$$

Rearranging this equation, we obtain

$$x^2 = \frac{y+1}{3}.$$

Since $y \in [-1, \infty)$, we know $y + 1 \geq 0$, so

$$x = \sqrt{\frac{y+1}{3}}$$

is in the domain $[0, \infty)$. So we have

$$\begin{aligned} f(x) &= 3x^2 - 1 \\ &= 3\left(\sqrt{\frac{y+1}{3}}\right)^2 - 1 \\ &= \frac{3(y+1)}{3} - 1 \\ &= (y+1) - 1 \\ &= y, \end{aligned}$$

that is, for every $y \in [-1, \infty)$ there is an $x \in [0, \infty)$ such that $f(x) = y$.

Thus $f([0,\infty)) \supseteq [-1,\infty)$.

Since $f([0,\infty)) \subseteq [-1,\infty)$ and $f([0,\infty)) \supseteq [-1,\infty)$, it follows that $f([0,\infty)) = [-1,\infty)$, so the image set of f is indeed $[-1,\infty)$.

The image set of f is equal to the codomain, so f is onto.

Under f, we know that y is the image of $\sqrt{(y+1)/3}$, so under the inverse, $\sqrt{(y+1)/3}$ is the image of y.

So f^{-1} is the function

$$\begin{aligned} f^{-1} : [-1,\infty) &\longrightarrow [0,\infty) \\ y &\longmapsto \sqrt{(y+1)/3}. \end{aligned}$$

This can be expressed in terms of x as

$$\begin{aligned} f^{-1} : [-1,\infty) &\longrightarrow [0,\infty) \\ x &\longmapsto \sqrt{(x+1)/3}. \end{aligned}$$

Exercise A33

For each of the following functions, determine whether f has an inverse function f^{-1} and, if it exists, find it.

(a) $\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 1 + x^2 \end{aligned}$

(b) $\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x,y) &\longmapsto (x,-y) \end{aligned}$

(c) $\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 8x + 3 \end{aligned}$

(For parts (a) and (b), use your answers from Exercises A30 and A32.)

Restrictions

When we are working with a function $f : A \longrightarrow B$, it is sometimes convenient to restrict attention to the behaviour of f on some subset C of A. For example, consider the function

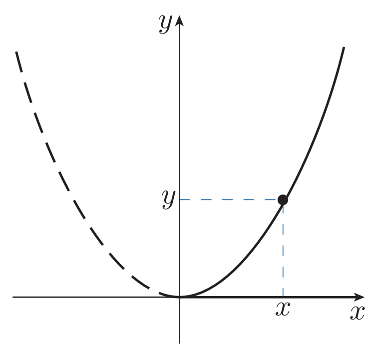
$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2. \end{aligned}$$

This function is not one-to-one and so does not have an inverse function. However, if the domain of f is replaced by the set $C = [0,\infty)$, then we obtain a related function,

$$\begin{aligned} g : C &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2, \end{aligned}$$

shown in Figure 45. The rule is the same as for f, but the domain is 'restricted' to produce a new function g that is one-to-one and so has an inverse.

The function g is an example of a *restriction* of f in the sense that $g(x) = f(x)$ for all x in the domain of g.



$y = x^2$

Figure 45 The function g with domain $[0,\infty)$

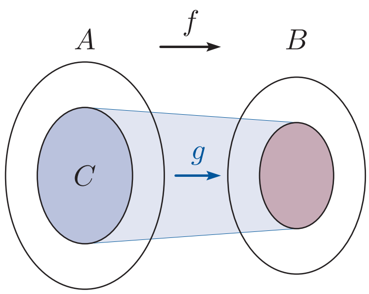


Figure 46 The function g is the restriction of f to C

More generally, we define a restriction, illustrated in Figure 46, as follows.

Definition

Let $f : A \longrightarrow B$ and let C be a subset of the domain A. Then the function $g : C \longrightarrow B$ defined by

$$g(x) = f(x), \quad \text{for } x \in C,$$

is the **restriction** of f to C.

Exercise A34

Let f be the function

$$\begin{aligned} f : \mathbb{R} &\longrightarrow [-1, 1] \\ x &\longmapsto \sin x. \end{aligned}$$

Write down a restriction of f that is one-to-one.

3.4 Composite functions

In Subsection 3.1, you saw how a function may be regarded as a machine that processes elements in the domain to produce elements in the codomain. Now suppose that two such machines are linked together, so that the elements emerging from the first machine are fed into the second machine for further processing. The overall effect is to create a new 'composite' machine that corresponds to a so-called *composite* function.

For example, consider the real functions

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 2x - 5. \end{aligned}$$

When the machines for f and g are linked together so that elements are first processed by f and then by g, we obtain the 'composite' machine illustrated by the large box in Figure 47.

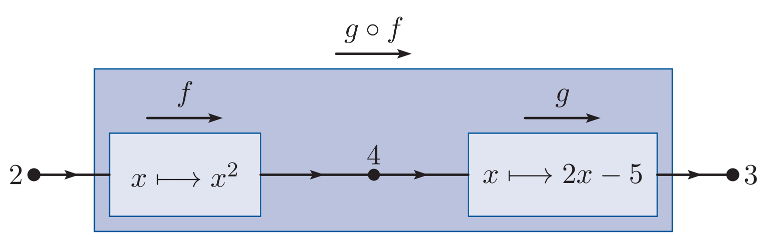


Figure 47 The composite function $g \circ f$ as a machine

For instance, when 2 is fed into the machine, it is first squared by f to produce the number 4, and then 4 is processed by g to give the number $(2 \times 4) - 5 = 3$.

Similarly, when an arbitrary real number x is fed into the machine, it is first processed by f to give the real number x^2. Since x^2 lies in $\mathbb{R}$, the domain of g, the number x^2 can then be processed by g to give $2x^2 - 5$. Thus, overall, the composite machine corresponds to a function, which we denote by $g \circ f$, whose rule is

$$(g \circ f)(x) = g(f(x)) = g(x^2) = 2x^2 - 5.$$

In general, we have the following definition.

Definition

Let $f : A \longrightarrow B$ and $g : B \longrightarrow C$ be two functions such that the domain of g is the same set as the codomain, B, of f. Then the **composite function** $g \circ f$ is given by

$$\begin{aligned} g \circ f : A &\longrightarrow C \\ x &\longmapsto g(f(x)). \end{aligned}$$

Notice that $g \circ f$ means f first, then g.

Exercise A35

Let f and g be the functions

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto -x \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 3x + 1. \end{aligned}$$

Determine the composite functions

(a) $g \circ f$, (b) $f \circ g$.

In general, the composite functions $g \circ f$ and $f \circ g$ are not equal, as you saw in Exercise A35.

Composite functions have many uses in mathematics; for example, we can use them to examine the effect of one transformation of the plane followed by another.

Suppose, for instance, that f and g are the reflections of the plane in the x-axis and y-axis respectively:

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x, -y) \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (-x, y). \end{aligned}$$

The composite function $g \circ f$ describes the overall effect of first reflecting in the x-axis (changing the sign of y) and then reflecting in the y-axis (changing the sign of x), as shown in Figure 48. The rule of $g \circ f$ is

$$\begin{aligned} (g \circ f)(x, y) = g(f(x, y)) &= g(x, -y) \\ &= (-x, -y). \end{aligned}$$

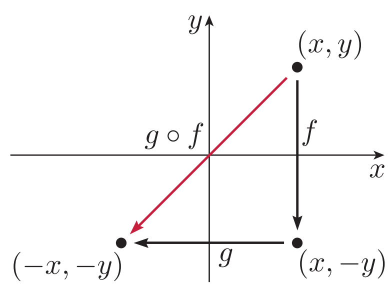


Figure 48 The composite $g \circ f$

Thus $g \circ f$ is the function

$$\begin{aligned} g \circ f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (-x, -y), \end{aligned}$$

which rotates the plane through an angle π about the origin, as can be seen by considering Figure 49, which shows how a square is transformed by $g \circ f$.

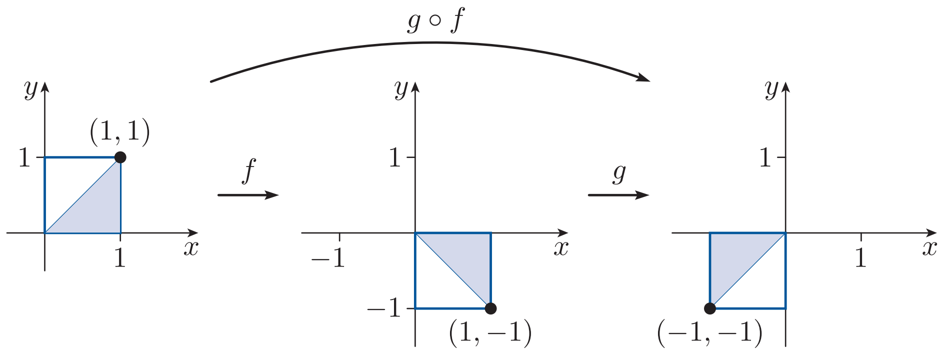


Figure 49 The composite function $g \circ f$ transforming a square

Exercise A36

Determine the composite function $f \circ g$, where f and g are the reflections of the plane in the x-axis and y-axis respectively, as defined above.

So far, we have considered the composite function $g \circ f$ only when the domain of the function g is the same as the codomain of the function f. We can, however, form the composite function $g \circ f$ when g and f are *any* two functions.

For example, consider the functions

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R} - \{1\} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{x-1}. \end{aligned}$$

Recall that $\mathbb{R} - \{1\}$ is the set of all real numbers with 1 excluded.

Here the domain of g is not equal to the codomain of f, but we can still consider the composite function $g \circ f$, with the rule

$$(g \circ f)(x) = g(f(x)) = g(x^2) = \frac{1}{x^2 - 1}.$$

However, we have to be careful about the domain of $g \circ f$. It cannot be the whole of $\mathbb{R}$, the domain of f. To see this, consider what happens when we try to feed the number 1 into the 'machine' corresponding to $g \circ f$, as shown in Figure 50.

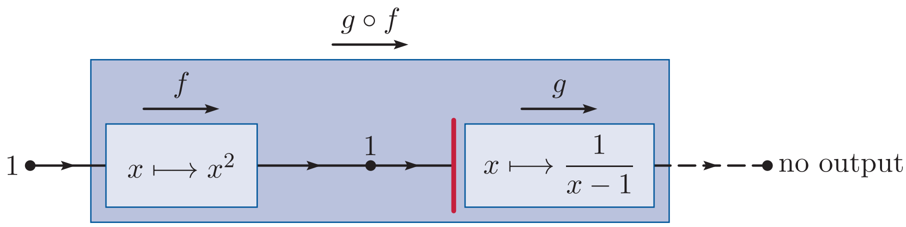


Figure 50 An input number that cannot be 'processed' by $g \circ f$

If we try to feed the number 1 into the machine, then it can be processed by f to produce the number 1, but 1 cannot then be processed by g, since it is not in the domain of g. We have the same problem if we try to feed the number -1 into the machine. However, if we feed any other number in the domain of f into the machine, then it can be processed by f and then g to produce a final output number. So we take the domain of $g \circ f$ to be $\mathbb{R} - \{1, -1\}$. Thus the composite function $g \circ f$ is

$$\begin{aligned} g \circ f : \mathbb{R} - \{1,-1\} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{x^2 - 1}. \end{aligned}$$

In general, if f and g are any two functions, then we take the domain of the composite function $g \circ f$ to consist of all the elements in the domain of f such that $f(x)$ is in the domain of g. The codomain of $g \circ f$ is always the same as the codomain of g. So we have the following definition.

Definition

Let $f : A \longrightarrow B$ and $g : C \longrightarrow D$ be any two functions; then the **composite function** $g \circ f$ has:

- domain $\{x \in A : f(x) \in C\}$
- codomain D
- rule $(g \circ f)(x) = g(f(x))$.

This definition allows us to consider the composite of *any* two functions, though in some cases the domain may turn out to be the empty set $\varnothing$. However, some texts insist on $f(A) \subseteq C$ as a condition to ensure $g \circ f$ exists.

In the example above with

$$\begin{aligned} g \circ f : \mathbb{R} - \{1, -1\} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{x^2 - 1}, \end{aligned}$$

the domain of $g \circ f$ is just the set of values for which the rule of $g \circ f$ is defined. This is not always the case, as illustrated in the following worked exercise where the domain of f is not the whole of $\mathbb{R}$.

Worked Exercise A15

Determine the composite function $g \circ f$ for the following functions f and g:

$$\begin{aligned} f : [0, 2\pi) &\longrightarrow [-1, 1] \\ x &\longmapsto \sin x \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R} - \{-1\} &\longrightarrow \mathbb{R}^* \\ x &\longmapsto \frac{1}{x+1}. \end{aligned}$$

Solution

The composite function $g \circ f$ means f then g.

The rule of $g \circ f$ is

$$(g \circ f)(x) = g(f(x)) = g(\sin x) = \frac{1}{\sin x + 1}.$$

A number x is in the domain of $g \circ f$ if it is in the domain of f and $f(x)$ is in the domain of g.

The domain of $g \circ f$ is

$$\{x \in [0, 2\pi) : f(x) \in \mathbb{R} - \{-1\}\}.$$

If $x \in [0, 2\pi)$, then $f(x) \in \mathbb{R} - \{-1\}$ unless $f(x) = -1$.

Now $f(x) = -1$ means $\sin x = -1$, and the only value of x in $[0, 2\pi)$ such that $\sin x = -1$ is

$$x = \frac{3\pi}{2}.$$

The domain is complicated to write down so it helps to give it a name, say D.

So the domain of $g \circ f$ is

$$D = [0, 2\pi) - \{3\pi/2\}.$$

Thus $g \circ f$ is the function

$$\begin{aligned} g \circ f : D &\longrightarrow \mathbb{R}^* \\ x &\longmapsto \frac{1}{\sin x + 1}. \end{aligned}$$

Notice that, as claimed, in the worked exercise above the domain of $g \circ f$ is not the full set of values for which $g \circ f$ is defined. The full set of values for which $g \circ f$ is defined is

$$\{x \in \mathbb{R} : \sin x \neq -1\} = \mathbb{R} - \left\{\left(2n - \tfrac{1}{2}\right)\pi : n \in \mathbb{Z}\right\}.$$

Exercise A37

Determine the composite function $g \circ f$ for the following functions f and g:

$$\begin{aligned} f : [-1,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto 3x+1 \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R} - \{-2\} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{3}{x+2}. \end{aligned}$$

Using function composition to show that a function is the inverse of another function

Suppose that $f : A \longrightarrow B$ is a one-to-one and onto function. Then f has an inverse function $f^{-1} : B \longrightarrow A$. We can therefore consider the effect that the composite function $f^{-1} \circ f : A \longrightarrow A$ has on an arbitrary element x in A. First, f maps x to an element $y = f(x)$ in B. Then f^{-1} 'undoes' the effect of f and maps y back to x, as illustrated in Figure 51. Overall, the effect of $f^{-1} \circ f$ is to leave x unchanged, or *fixed*: that is, $(f^{-1} \circ f)(x) = x$. Since x is an arbitrary element of A, it follows that $f^{-1} \circ f$ fixes all the elements of A. In other words, $f^{-1} \circ f = i_A$, the identity function on set A.

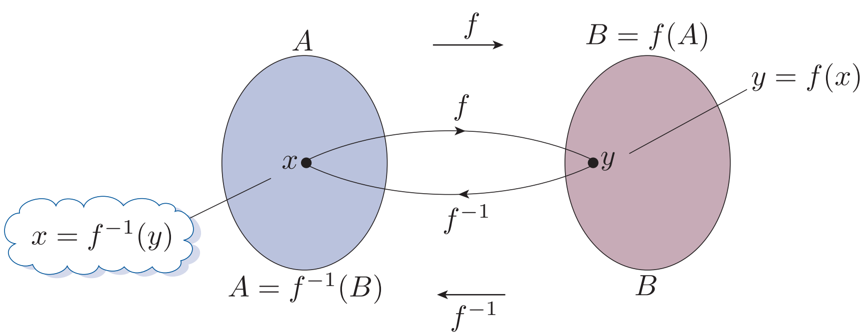


Figure 51 The composite function $f^{-1} \circ f$

A similar argument can be used to show that $f \circ f^{-1} = i_B$. So, if $f : A \longrightarrow B$ has an inverse function $f^{-1} : B \longrightarrow A$, then

$$f^{-1} \circ f = i_A \quad \text{and} \quad f \circ f^{-1} = i_B.$$

The *converse* of this statement is also true: that is, if a function $g : B \longrightarrow A$ satisfies

$$g \circ f = i_A \quad \text{and} \quad f \circ g = i_B,$$

then g is the inverse function of f. A proof of this is given after Exercise A39. It leads to the following strategy.

Strategy A2

To show that the function $g : B \longrightarrow A$ is the inverse function of the function $f : A \longrightarrow B$:

1. show that $g(f(x)) = x$ for each $x \in A$; that is, $g \circ f = i_A$
2. show that $f(g(y)) = y$ for each $y \in B$; that is, $f \circ g = i_B$.

In practice, we can sometimes use Strategy A2 as an alternative way of *finding* an inverse function. We make an inspired guess at the inverse function, and use Strategy A2 to check that our guess is correct.

Worked Exercise A16

Use Strategy A2 to find the inverse of the function

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \tfrac{1}{2}x. \end{aligned}$$

Solution

We guess that the inverse function is $g(x) = 2x$.

Let

$$\begin{aligned} g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 2x. \end{aligned}$$

We use Strategy A2 to check that our guess is correct.

The domain of f is $\mathbb{R}$, and for each $x \in \mathbb{R}$ we have

$$g(f(x)) = g\left(\tfrac{1}{2}x\right) = 2 \times \tfrac{1}{2}x = x;$$

that is, $g \circ f = i_{\mathbb{R}}$.

The domain of g is also $\mathbb{R}$, and for each $y \in \mathbb{R}$ we have

$$f(g(y)) = f(2y) = \tfrac{1}{2} \times 2y = y;$$

that is, $f \circ g = i_{\mathbb{R}}$.

Since $g \circ f = i_{\mathbb{R}}$ and $f \circ g = i_{\mathbb{R}}$, it follows that g is the inverse function of f.

Exercise A38

Use Strategy A2 to show that g is the inverse of f, where

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 5x - 3 \end{aligned} \qquad \text{and} \qquad \begin{aligned} g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{x+3}{5}. \end{aligned}$$

Exercise A39

Use Strategy A2 to find the inverse of the function

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x - 1, y + 3). \end{aligned}$$

To end this section, here is the promised proof that if the functions $f : A \longrightarrow B$ and $g : B \longrightarrow A$ satisfy

$$g \circ f = i_A \quad \text{and} \quad f \circ g = i_B,$$

then g is the inverse function of f. That is, we prove that if the two steps of Strategy A2 hold, then f has an inverse function, and the inverse function is equal to g.

Suppose, then, that the two steps of Strategy A2 hold. First we show that f is one-to-one.

Suppose that $f(x_1) = f(x_2)$; then

$$g(f(x_1)) = g(f(x_2)),$$

so, since $g(f(x)) = x$ for each $x \in A$ by the first step of Strategy A2, we have $x_1 = x_2$. Thus f is one-to-one and so it has an inverse function f^{-1}.

Now we find the image set of f.

We know that the image set of f is a subset of its codomain B, so $f(A) \subseteq B$. We now show that $f(A) \supseteq B$ by showing that every element y of B is the image under f of some element in A. Suppose that $y \in B$. Then, by the second step of Strategy A2,

$$f(g(y)) = y;$$

that is, y is the image under f of the element $g(y)$ and $g(y) \in A$, as required. Therefore $f(A) \supseteq B$.

Since $f(A) \subseteq B$ and $f(A) \supseteq B$, it follows that the image set of f is B (that is, f is onto), and so f^{-1} has domain B.

We now know that each of the functions f^{-1} and g has domain B and codomain A. To show that they are equal, it remains to show that $g(y) = f^{-1}(y)$ for each element y of B.

Let y be an arbitrary element of B. Then $y = f(x)$ for some element x of A. So

$$f^{-1}(y) = x,$$

and, by the first step of Strategy A2,

$$g(y) = g(f(x)) = x.$$

Hence f^{-1} and g are indeed equal functions.

4 Vectors

In this section you will revise vectors, in both the plane $\mathbb{R}^2$ and in three-dimensional space $\mathbb{R}^3$. Vectors are used throughout Book C *Linear algebra*.

4.1 What is a vector?

A mathematical or physical quantity that has a direction as well as a size is called a **vector**, or a **vector quantity**. An example of such a quantity is *velocity*: to state the velocity of a car you have to give its speed and also the direction in which it is moving. In contrast, some mathematical and physical quantities, such as temperature and volume, have only a size – they have no direction associated with them. We call such quantities **scalars**, or **scalar quantities**. When discussing vectors and scalars, we usually use the term **magnitude**, rather than size.

Definition

A **vector** is a quantity that is determined by its magnitude and direction. A **scalar** is a quantity that is determined by its magnitude.

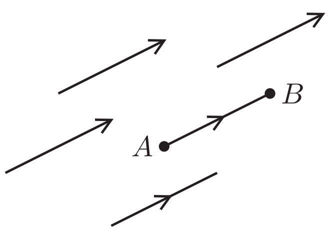


Figure 52 The same vector represented in different ways

We can represent a vector in $\mathbb{R}^2$ or in $\mathbb{R}^3$ geometrically by a line segment with an arrowhead, as illustrated in Figure 52. The length of the line segment is a measure of the magnitude of the vector, and the direction of the arrowhead indicates the direction. The starting point of the line segment does not matter; for example, all the line segments with arrowheads in Figure 52 represent the same vector. We can draw the arrowhead at the end of the line segment, or in the middle of it, as convenient. A vector represented by a line segment from A to B, with an arrowhead pointing from A to B, can be written as $\overrightarrow{AB}$.

Often we use single letters, such as $\mathbf{a}$, $\mathbf{b}$, $\mathbf{p}$, $\mathbf{q}$ or $\mathbf{v}$, to denote vectors. Vectors are usually distinguished in print by the use of a bold typeface, and in handwritten work by underlining the letters (for example, $\underline{v}$). These are important conventions as they clearly distinguish vector quantities from scalar quantities.

We denote the magnitude of a vector $\mathbf{v}$ by the notation $|\mathbf{v}|$.

There is one vector that does not fit conveniently into the definition above; namely, the *zero vector*. It represents any vector quantity that has magnitude zero and hence has no direction, such as the velocity of a stationary car.

Definition

The **zero vector** is the vector whose magnitude is zero, and whose direction is undefined. It is denoted by the symbol $\mathbf{0}$.

The next box defines what it means to say that two vectors are equal.

Definition

Two vectors $\mathbf{a}$ and $\mathbf{b}$ are **equal** if:

- they have the same magnitude; that is, $|\mathbf{a}| = |\mathbf{b}|$
- they are in the same direction.

We write $\mathbf{a} = \mathbf{b}$.

For example, in Figure 53, the vector $\mathbf{v}$ is equal to the vector $\mathbf{d}$, but is not equal to any of the other vectors, as they all differ from $\mathbf{v}$ in magnitude or direction.

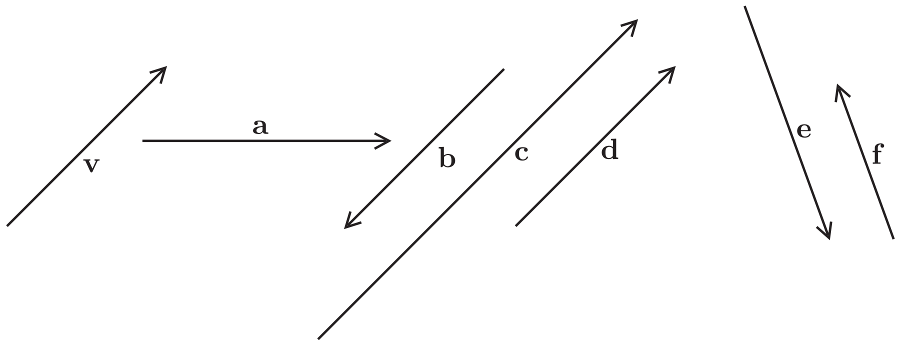


Figure 53 A selection of vectors in the plane

We now briefly revise some other definitions relating to vectors.

Definition

The **negative** of a vector $\mathbf{v}$ is the vector that has the same magnitude as $\mathbf{v}$, but the opposite direction. It is denoted by $-\mathbf{v}$.

For example, in Figure 53 we have $\mathbf{b} = -\mathbf{v}$. If we write $\mathbf{v}$ as $\overrightarrow{AB}$ for suitable points A and B, then $-\mathbf{v} = \overrightarrow{BA}$, as shown in Figure 54.

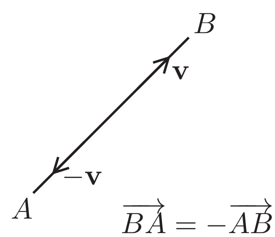


Figure 54 The vectors $\mathbf{v}$ and $-\mathbf{v}$

Scalar multiple of a vector

Let k be a scalar and $\mathbf{v}$ a vector. The **scalar multiple** $k\mathbf{v}$ of $\mathbf{v}$ is the vector:

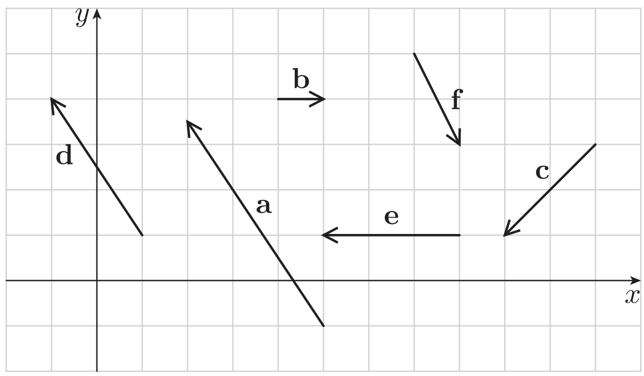
- whose magnitude is $|k|$ times the magnitude of $\mathbf{v}$; that is, $|k\mathbf{v}| = |k|\,|\mathbf{v}|$
- that has the same direction as $\mathbf{v}$ if $k > 0$, and the opposite direction if $k < 0$.

If $k = 0$, then $k\mathbf{v} = \mathbf{0}$.

For example, in Figure 53 we have $\mathbf{c} = 2\mathbf{v}$, since $\mathbf{c}$ has the same direction as $\mathbf{v}$ but twice the magnitude, and $\mathbf{e} = -\frac{3}{2}\mathbf{f}$, since $\mathbf{e}$ has the opposite direction to $\mathbf{f}$ and its magnitude is $\frac{3}{2}$ times that of $\mathbf{f}$.

Exercise A40

For each of the vectors shown below, decide whether it is a multiple of any of the other vectors; if it is, write down an equation of the form $\mathbf{v}_1 = k\mathbf{v}_2$ that specifies the relationship between them.



Exercise A41

For the vector $\mathbf{d}$ in Exercise A40, sketch $3\mathbf{d}$ and $-2\mathbf{d}$.

We can add two vectors using either of the two laws below. They give the same result, as illustrated in Figure 55.

Triangle Law for addition of vectors

The sum $\mathbf{p} + \mathbf{q}$ of two vectors $\mathbf{p}$ and $\mathbf{q}$ is obtained as follows.

1. Starting at any point, draw the vector $\mathbf{p}$.
2. Starting from the tip of the vector $\mathbf{p}$, draw the vector $\mathbf{q}$.

Then the sum $\mathbf{p} + \mathbf{q}$ is the vector from the tail of $\mathbf{p}$ to the tip of $\mathbf{q}$.

Parallelogram Law for addition of vectors

The sum $\mathbf{p} + \mathbf{q}$ of two vectors $\mathbf{p}$ and $\mathbf{q}$ is obtained as follows.

1. Starting at the same point, draw the vectors $\mathbf{p}$ and $\mathbf{q}$.
2. Complete the parallelogram of which these vectors are adjacent sides.

Then the sum $\mathbf{p} + \mathbf{q}$ is the vector from the point where the tails of $\mathbf{p}$ and $\mathbf{q}$ meet to the opposite corner of the parallelogram.

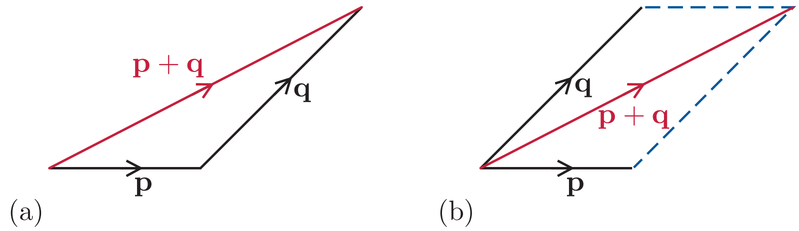


Figure 55 The sum $\mathbf{p} + \mathbf{q}$ obtained by (a) the Triangle Law (b) the Parallelogram Law

Addition and scalar multiplication of vectors obey the usual rules of algebra. The most important of these are listed in the box below.

Properties of vector algebra

Let $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ be vectors, and let $a, b \in \mathbb{R}$. The following properties hold.

Commutativity $\quad \mathbf{p} + \mathbf{q} = \mathbf{q} + \mathbf{p}$

Associativity $\quad (\mathbf{p} + \mathbf{q}) + \mathbf{r} = \mathbf{p} + (\mathbf{q} + \mathbf{r})$

Distributivity $\quad a(\mathbf{p} + \mathbf{q}) = a\mathbf{p} + a\mathbf{q},$

$\quad (a + b)\mathbf{p} = a\mathbf{p} + b\mathbf{p}.$

Finally, we define subtraction of vectors in terms of addition and the negative of a vector, as follows, and as illustrated in Figure 56.

Definition

The **difference** $\mathbf{p} - \mathbf{q}$ of the vectors $\mathbf{p}$ and $\mathbf{q}$ is

$$\mathbf{p} - \mathbf{q} = \mathbf{p} + (-\mathbf{q}).$$

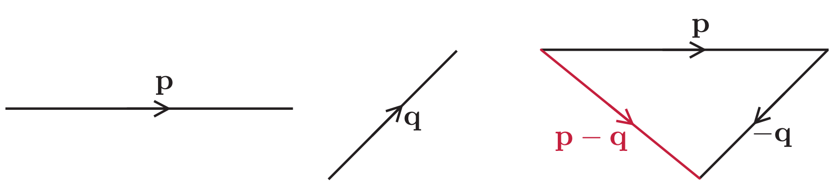


Figure 56 The difference $\mathbf{p} - \mathbf{q}$ of vectors $\mathbf{p}$ and $\mathbf{q}$

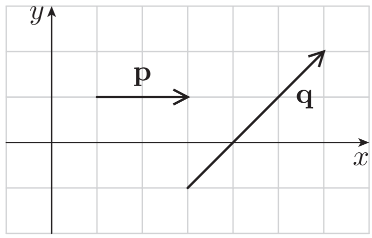
Since the vector $-\mathbf{q}$ has the same magnitude as $\mathbf{q}$ but the opposite direction, we can draw $\mathbf{p} - \mathbf{q}$ by using either of the two constructions that we use for adding vectors.

In general, $\mathbf{q} - \mathbf{p}$ does not equal $\mathbf{p} - \mathbf{q}$; in fact, as you would expect,

$$\mathbf{q} - \mathbf{p} = -(\mathbf{p} - \mathbf{q}).$$

Exercise A42

For the vectors $\mathbf{p}$ and $\mathbf{q}$ shown below, sketch $\mathbf{p}+\mathbf{q}$, $\mathbf{p}-\mathbf{q}$ and $2\mathbf{p}+\frac{1}{2}\mathbf{q}$.



4.2 Components and the arithmetic of vectors

We can sometimes simplify the manipulation of vectors by expressing them in *component form*. To do this, we start by defining the following *unit vectors*, shown in Figure 57. A **unit vector** is a vector of magnitude 1.

In $\mathbb{R}^2$, the vectors $\mathbf{i}$ and $\mathbf{j}$ are the unit vectors in the positive directions of the x- and y-axes, respectively.

In $\mathbb{R}^3$, the vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are the unit vectors in the positive directions of the x-, y- and z-axes, respectively.

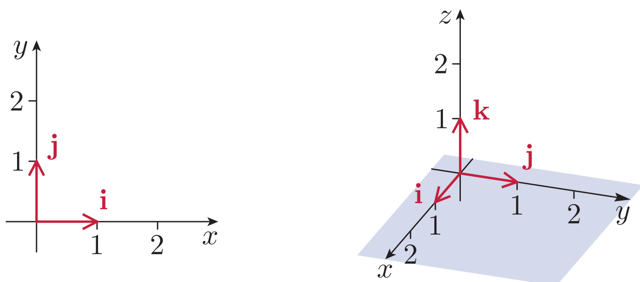


Figure 57 The unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$

Any vector in $\mathbb{R}^2$ can be expressed as the sum of scalar multiples of $\mathbf{i}$ and $\mathbf{j}$, and similarly any vector in $\mathbb{R}^3$ can be expressed as the sum of scalar multiples of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. For example, the vector $\mathbf{v}$ in Figure 58(a) can be expressed as

$$\mathbf{v} = 3\mathbf{i} + 4\mathbf{j},$$

and the vector $\mathbf{w}$ in Figure 58(b) can be expressed as

$$\mathbf{w} = 2\mathbf{i} + 4\mathbf{j} + 3\mathbf{k}.$$

These expressions are the *component forms* of $\mathbf{v}$ and $\mathbf{w}$.

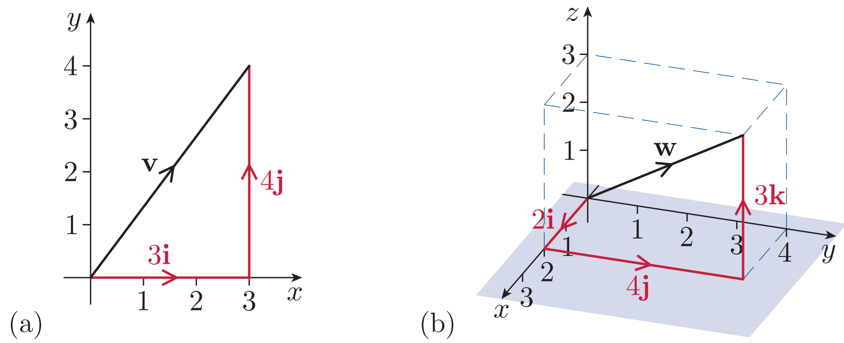


Figure 58 (a) A vector $\mathbf{v}$ in $\mathbb{R}^2$ (b) A vector $\mathbf{w}$ in $\mathbb{R}^3$

In general we have the following.

Definitions

Any vector $\mathbf{p}$ in $\mathbb{R}^2$ can be expressed in **component form** as

$$\mathbf{p} = a_1\mathbf{i} + a_2\mathbf{j}, \quad \text{for some real numbers } a_1, a_2;$$

we often write $\mathbf{p} = (a_1, a_2)$, for brevity. The numbers a_1 and a_2 are the **components** of $\mathbf{p}$ in the x- and y-directions, respectively.

Any vector $\mathbf{p}$ in $\mathbb{R}^3$ can be expressed in **component form** as

$$\mathbf{p} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}, \quad \text{for some real numbers } a_1, a_2, a_3;$$

we often write $\mathbf{p} = (a_1, a_2, a_3)$, for brevity. The numbers a_1, a_2 and a_3 are the **components** of $\mathbf{p}$ in the x-, y- and z-directions, respectively.

So, for example, the component form of the vector $\mathbf{v}$ in Figure 58(a) is

$$3\mathbf{i} + 4\mathbf{j}, \quad \text{or, equivalently,} \quad (3, 4).$$

Similarly, the component form of the vector $\mathbf{w}$ in Figure 58(b) is

$$2\mathbf{i} + 4\mathbf{j} + 3\mathbf{k}, \quad \text{or, equivalently,} \quad (2, 4, 3).$$

In some texts, the ordered pairs and ordered triples that represent the component forms of vectors are written vertically, as

$$\begin{pmatrix} 3 \\ 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 4 \\ 3 \end{pmatrix},$$

for example, to distinguish them from points. Although we write them horizontally in this module, the meaning of an ordered pair or ordered triple should be clear from the context.

Exercise A43

Sketch the following vectors in $\mathbb{R}^2$ on a single diagram:

$$2\mathbf{i} - 3\mathbf{j}, \quad -3\mathbf{i} + 4\mathbf{j}, \quad -2\mathbf{i} - 2\mathbf{j}.$$

In the box below, the operations on vectors that were described geometrically in Subsection 4.1 are expressed in terms of components. The component forms of the vectors are expressed as ordered pairs and ordered triples in the box; there are analogous formulas for vectors expressed in terms of the unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. For example, the zero vector in $\mathbb{R}^2$ can be written as $\mathbf{0} = 0\mathbf{i} + 0\mathbf{j}$ rather than as $\mathbf{0} = (0, 0)$.

Vector arithmetic in component form

Equality Two vectors, both in $\mathbb{R}^2$ or both in $\mathbb{R}^3$, are equal if their corresponding components are equal.

Zero vector The zero vector is

$$\mathbf{0} = (0, 0) \quad \text{in } \mathbb{R}^2,$$

$$\mathbf{0} = (0, 0, 0) \quad \text{in } \mathbb{R}^3.$$

Addition To add vectors in $\mathbb{R}^2$ or in $\mathbb{R}^3$, add their corresponding components:

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2),$$

$$(a_1, a_2, a_3) + (b_1, b_2, b_3) = (a_1 + b_1, a_2 + b_2, a_3 + b_3).$$

Negatives To find the negative of a vector in $\mathbb{R}^2$ or in $\mathbb{R}^3$, take the negatives of its components:

$$-(a_1, a_2) = (-a_1, -a_2),$$

$$-(a_1, a_2, a_3) = (-a_1, -a_2, -a_3).$$

Subtraction To subtract vectors in $\mathbb{R}^2$ or in $\mathbb{R}^3$, subtract the corresponding components:

$$(a_1, a_2) - (b_1, b_2) = (a_1 - b_1, a_2 - b_2),$$

$$(a_1, a_2, a_3) - (b_1, b_2, b_3) = (a_1 - b_1, a_2 - b_2, a_3 - b_3).$$

Scalar multiplication To multiply a vector in $\mathbb{R}^2$ or in $\mathbb{R}^3$ by a real number k, multiply each component by k:

$$k(a_1, a_2) = (ka_1, ka_2),$$

$$k(a_1, a_2, a_3) = (ka_1, ka_2, ka_3).$$

Magnitude The magnitude of the vector (a_1, a_2) in $\mathbb{R}^2$ is

$$\sqrt{a_1^2 + a_2^2}.$$

The magnitude of the vector (a_1, a_2, a_3) in $\mathbb{R}^3$ is

$$\sqrt{a_1^2 + a_2^2 + a_3^2}.$$

The formulas for magnitude are derived from the distance formulas for $\mathbb{R}^2$ and $\mathbb{R}^3$ that you met in Section 1.

Here are some examples of vector arithmetic in component form, in $\mathbb{R}^2$: the sum of two vectors,

$$(1,-3)+(4,2)=(1+4,-3+2)=(5,-1),$$

the negative of a vector,

$$-(1,-3)=(-1,3),$$

and a scalar multiple of a vector,

$$2(2,-1)=(4,-2).$$

The magnitude of the vector $(1,-3)$ is given by

$$\sqrt{1^2+(-3)^2}=\sqrt{1+9}=\sqrt{10}.$$

Exercise A44

For each of the following pairs of vectors $\mathbf{p}$ and $\mathbf{q}$, write down $\mathbf{p}+\mathbf{q}$, $-\mathbf{q}$ and $\mathbf{p}-\mathbf{q}$.

(a) $\mathbf{p}=(3,-1)$ and $\mathbf{q}=(-1,-2)$.

(b) $\mathbf{p}=-\mathbf{i}-2\mathbf{j}$ and $\mathbf{q}=2\mathbf{i}-\mathbf{j}$.

(c) $\mathbf{p}=-\mathbf{i}+2\mathbf{k}$ and $\mathbf{q}=\mathbf{i}-2\mathbf{j}-\mathbf{k}$.

Exercise A45

For each of the following pairs of vectors $\mathbf{p}$ and $\mathbf{q}$, determine $2\mathbf{p}$, $3\mathbf{q}$ and $2\mathbf{p}-3\mathbf{q}$, and find the magnitude of $\mathbf{q}$.

(a) $\mathbf{p}=(3,-1)$ and $\mathbf{q}=(-1,-2)$.

(b) $\mathbf{p}=-\mathbf{i}+2\mathbf{k}$ and $\mathbf{q}=\mathbf{i}-2\mathbf{j}-\mathbf{k}$.

Unit vectors

As you saw earlier, a **unit vector** is a vector of magnitude 1. We denote the unit vector that is in the same direction as a particular vector $\mathbf{v}$ by $\widehat{\mathbf{v}}$ (read as 'v hat'), as illustrated in Figure 59.

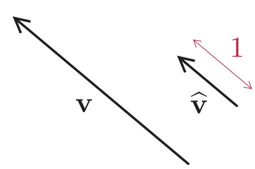


Figure 59 A vector $\mathbf{v}$ and its corresponding unit vector $\widehat{\mathbf{v}}$

To find $\widehat{\mathbf{v}}$, we multiply $\mathbf{v}$ by the reciprocal of its magnitude, as follows.

The unit vector in the same direction as a vector $\mathbf{v}$ is

$$\widehat{\mathbf{v}}=\frac{1}{|\mathbf{v}|}\mathbf{v}.$$

The exception to this notation for unit vectors is that we use the special symbols $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ for the unit vectors in the positive directions of the x-, y- and z-axes, as you saw earlier. This is common practice, though some texts use the alternative symbols $\widehat{\mathbf{x}}$, $\widehat{\mathbf{y}}$ and $\widehat{\mathbf{z}}$ for these vectors.

Worked Exercise A17

Find $\widehat{\mathbf{v}}$ for $\mathbf{v} = (3, 4)$.

Solution

For $\mathbf{v} = (3, 4)$ we have

$$|\mathbf{v}| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5,$$

so

$$\widehat{\mathbf{v}} = \frac{1}{5}(3, 4) = \left(\frac{3}{5}, \frac{4}{5}\right).$$

Exercise A46

Find $\widehat{\mathbf{v}}$ for each of the following vectors $\mathbf{v}$.

(a) $(2, -3)$ (b) $5\mathbf{i} + 12\mathbf{j}$

Position vectors

There is a natural and useful way to associate every point in the plane or in three-dimensional space with a vector. We make the following definition.

Definition

Let P be any point in $\mathbb{R}^2$ or $\mathbb{R}^3$. The **position vector** of P is the vector whose starting point is the origin and whose finishing point is P, that is, the vector $\overrightarrow{OP}$, where O is the origin.

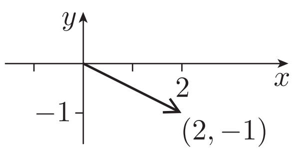


Figure 60 The position vector of the point $(2, -1)$

For example, the position vector of the point $P(2, -1)$ is the vector $\overrightarrow{OP} = 2\mathbf{i} - \mathbf{j}$ (often written as $(2, -1)$), as shown in Figure 60.

In general, any point (x, y) in $\mathbb{R}^2$ has position vector $x\mathbf{i} + y\mathbf{j}$ (often written as (x, y)), and similarly any point (x, y, z) in $\mathbb{R}^3$ has position vector $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ (often written as (x, y, z)).

Exercise A47

Let $\mathbf{p}$ and $\mathbf{q}$ be the position vectors of the points $(5, 3)$ and $(1, 4)$, respectively.

(a) Determine the vectors $\mathbf{p} - \mathbf{q}$, $\mathbf{p} + \mathbf{q}$ and $\frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{q}$.

(b) Sketch $\mathbf{p}$, $\mathbf{q}$ and each of the vectors that you found in part (a), starting each vector at the origin.

The following simple result about position vectors is often useful.

Let A and B be points (in $\mathbb{R}^2$ or $\mathbb{R}^3$), with position vectors $\mathbf{a}$ and $\mathbf{b}$, respectively. Then

$$\overrightarrow{AB} = \mathbf{b} - \mathbf{a}.$$

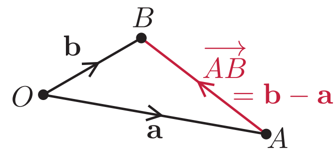


Figure 61 Points A and B and their position vectors

To see this, let O be the origin, as shown in Figure 61. Then

$$\begin{aligned}\overrightarrow{AB} &= \overrightarrow{AO} + \overrightarrow{OB} \quad \text{(by the Triangle Law for vector addition)}\\ &= -\overrightarrow{OA} + \overrightarrow{OB}\\ &= -\mathbf{a} + \mathbf{b}\\ &= \mathbf{b} - \mathbf{a},\end{aligned}$$

as claimed.

The sets $\mathbb{R}^2$ and $\mathbb{R}^3$

Finally, we clarify some issues about the sets $\mathbb{R}^2$ and $\mathbb{R}^3$. You have seen that we use the notation $\mathbb{R}^2$ to denote the plane, and the notation $\mathbb{R}^3$ to denote three-dimensional space. Strictly, the meaning of these notations is as follows:

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\},$$
$$\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}.$$

That is, $\mathbb{R}^2$ is the set of all ordered pairs of real numbers, and $\mathbb{R}^3$ is the set of all ordered triples of real numbers. We interpret these sets as the plane and as three-dimensional space, respectively, by interpreting their elements as the coordinates of points with respect to particular coordinate systems, in the way that you have seen.

However, it is often useful to instead interpret the elements of $\mathbb{R}^2$ and $\mathbb{R}^3$ as *vectors*. For example, we can interpret the element $(2, -1)$ of $\mathbb{R}^2$ not as the point with coordinates $(2, -1)$, but instead as the vector with component form $(2, -1)$.

We can use whichever interpretation of $\mathbb{R}^2$ and $\mathbb{R}^3$ is more useful in a particular context. A link between the two interpretations is provided by position vectors, because the vector with component form (x, y) is the position vector of the point with coordinates (x, y), and similarly the vector with component form (x, y, z) is the position vector of the point with coordinates (x, y, z).

This link also makes it straightforward to represent a particular point not by coordinates, but by a vector: we use its position vector. It might seem that this amounts to much the same thing, but the advantage of representing points by vectors is that it enables us to use the properties of vectors to work with points. This leads to some very convenient ways of working with points, as you will see in the next subsection and again in Subsection 4.5. In Book C you will see how generalising all these ideas leads to some interesting and very useful mathematics.

4.3 Vector form of the equation of a line

In Subsection 1.1, we found that every line in the plane has an equation of the form

$$ax + by = c,$$

where a, b and c are real numbers, with a and b not both zero. In this subsection we find an equivalent general form for the equation of a line in terms of vectors. Unlike the equation above, this vector form applies to lines in $\mathbb{R}^3$ as well as in $\mathbb{R}^2$, as you will see later in this subsection.

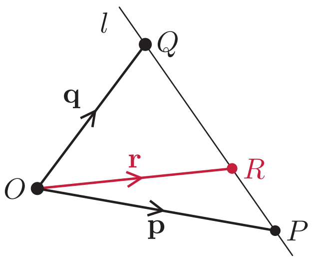


Figure 62 A point R on the line l

Let P and Q be points with position vectors $\mathbf{p}$ and $\mathbf{q}$, respectively, and let l be the line that passes through P and Q, as illustrated in Figure 62. We now find an expression for the position vector $\mathbf{r}$ of an arbitrary point R on l in terms of the position vectors $\mathbf{p}$ and $\mathbf{q}$.

Since the vector $\overrightarrow{PR}$ is parallel to the vector $\overrightarrow{PQ}$, it must be a multiple of $\overrightarrow{PQ}$, that is,

$$\overrightarrow{PR} = \lambda\overrightarrow{PQ},$$

for some real number λ. Now, by the result about position vectors given at the end of the last subsection, we have

$$\overrightarrow{PR} = \mathbf{r} - \mathbf{p} \quad \text{and} \quad \overrightarrow{PQ} = \mathbf{q} - \mathbf{p}.$$

So

$$\mathbf{r} - \mathbf{p} = \lambda(\mathbf{q} - \mathbf{p}).$$

We can rearrange this equation as

$$\mathbf{r} = \mathbf{p} + \lambda(\mathbf{q} - \mathbf{p}),$$

that is,

$$\mathbf{r} = (1 - \lambda)\mathbf{p} + \lambda\mathbf{q}. \qquad (3)$$

This is a general formula for the position vector of a point on the line through P and Q, in the following sense: each point on l corresponds to a particular value of λ, and vice versa. As shown in Figure 63, we have the following.

- If $\lambda = 0$, then $\mathbf{r} = 1\mathbf{p} + 0\mathbf{q} = \mathbf{p}$.
- If $\lambda = 1$, then $\mathbf{r} = 0\mathbf{p} + 1\mathbf{q} = \mathbf{q}$.
- If $\lambda > 1$, then R lies on l beyond Q.
- If $0 < \lambda < 1$, then R lies on l between P and Q.
- If $\lambda < 0$, then R lies on l beyond P.

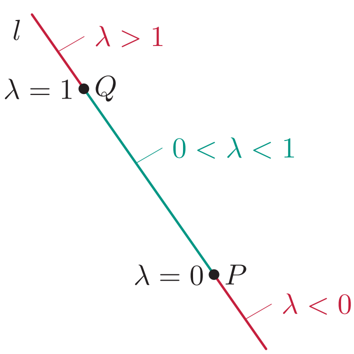


Figure 63 The position of R determined by λ

So we can regard equation (3) as the *vector form of the equation of the line* l.

Vector form of the equation of a line

The equation of the line through the points with position vectors $\mathbf{p}$ and $\mathbf{q}$ is

$$\mathbf{r} = (1-\lambda)\mathbf{p} + \lambda\mathbf{q}, \quad \text{where } \lambda \in \mathbb{R}.$$

Note in particular that when $\lambda = \frac{1}{2}$ in the equation above, we have $\mathbf{r} = \frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{q} = \frac{1}{2}(\mathbf{p}+\mathbf{q})$, which is the position vector of the *midpoint* of the line segment PQ.

Worked Exercise A18

(a) Let P and Q be the points with position vectors $\mathbf{p} = (1,3)$ and $\mathbf{q} = (-1,-2)$, respectively. Find the vector form of the equation of the line l through P and Q.

(b) Determine whether the point $(3,8)$ lies on l.

Solution

(a) The vector form of the equation of the line l is

$$\mathbf{r} = (1-\lambda)(1,3) + \lambda(-1,-2),$$

that is,

$$\mathbf{r} = (1-2\lambda, 3-5\lambda),$$

where $\lambda \in \mathbb{R}$.

(b) The point $(3,8)$ lies on the line if there is some real number λ such that

$$(3,8) = (1-2\lambda, 3-5\lambda).$$

Equating the corresponding components gives

$$3 = 1-2\lambda \quad \text{and} \quad 8 = 3-5\lambda.$$

The first equation gives $\lambda = -1$, and this value of λ also satisfies the other equation. Hence the point $(3,8)$ does lie on the line l.

Exercise A48

Let P and Q be the points with position vectors $\mathbf{p} = (3,1)$ and $\mathbf{q} = (2,3)$, respectively. Let l be the line through P and Q.

(a) Find the vector form of the equation of the line l.

(b) Determine the three points on l whose position vectors are given by the equation you found in part (a) when λ takes the values $\frac{2}{3}$, $\frac{3}{2}$ and $-\frac{1}{2}$, respectively.

(c) On a single diagram, sketch P, Q, the line l through P and Q, and the three points that you found in part (b).

Exercise A49

Let P, Q and l be as in Exercise A48.

(a) Determine the value of λ corresponding to the point $(4, -1)$ in the vector form of the equation of l.

(b) Use the vector form of the equation of l to prove that the point $\left(\frac{1}{2}, \frac{1}{2}\right)$ does not lie on l.

In the vector form of the equation of a line, there is no assumption that $\mathbf{p}$ and $\mathbf{q}$ are position vectors of points in $\mathbb{R}^2$: they may equally well be position vectors in $\mathbb{R}^3$.

Worked Exercise A19

(a) Let P and Q be the points with position vectors $\mathbf{p} = (1, 2, 3)$ and $\mathbf{q} = (3, -2, 1)$, respectively. Find the vector form of the equation of the line l through P and Q.

(b) Determine whether the point $(4, -4, 0)$ lies on the line l.

Solution

(a) The vector form of the equation of l is

$$\begin{aligned}\mathbf{r} &= (1-\lambda)(1,2,3) + \lambda(3,-2,1)\\ &= (1+2\lambda,\ 2-4\lambda,\ 3-2\lambda), \quad \text{where } \lambda \in \mathbb{R}.\end{aligned}$$

(b) The point $(4, -4, 0)$ lies on the line l if there is a real number λ such that

$$\begin{aligned}1 + 2\lambda &= 4,\\ 2 - 4\lambda &= -4,\\ 3 - 2\lambda &= 0.\end{aligned}$$

The first equation gives $\lambda = \frac{3}{2}$, and this value of λ also satisfies the other two equations. Hence the point $(4, -4, 0)$ lies on the line l.

Exercise A50

(a) Let P and Q be the points with position vectors $\mathbf{p} = (2, 1, 0)$ and $\mathbf{q} = (1, 0, -1)$, respectively. Find the vector form of the equation of the line l through P and Q.

(b) Determine the points on l whose position vectors are given by the equation you found in part (a) when λ takes the values $\frac{1}{2}$ and -1.

4.4 Scalar product

In this subsection you will meet a way of combining two vectors, known as the *scalar product* or *dot product*, which is useful in linear algebra, as you will see in Book C.

The definition of the scalar product is given below. It applies to vectors in both $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition

If $\mathbf{u}$ and $\mathbf{v}$ are non-zero vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, then the **scalar product** (or **dot product**) of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos\theta,$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

If one or both of $\mathbf{u}$ and $\mathbf{v}$ is the zero vector, then $\mathbf{u} \cdot \mathbf{v} = 0$.

The scalar product of two vectors is a scalar, hence the name.

Note that the angle between two vectors is defined to be the angle θ in the range $0 \leq \theta \leq \pi$ between their directions when the vectors are placed to have the same starting point (not necessarily the origin), as illustrated in Figure 64 for vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

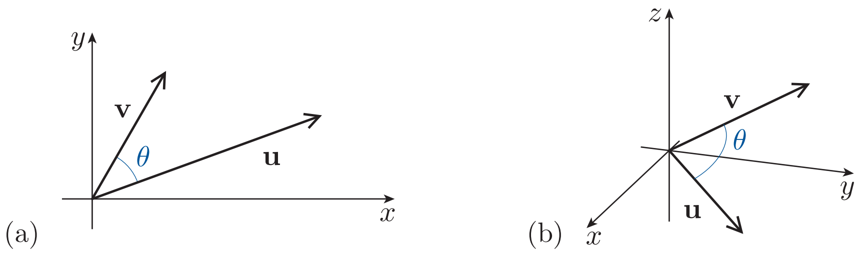


Figure 64 The angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ in (a) $\mathbb{R}^2$ and (b) $\mathbb{R}^3$

Let us use the definition of the scalar product to calculate the scalar product $\mathbf{u} \cdot \mathbf{v}$ of the vectors $\mathbf{u} = (2, 0)$ and $\mathbf{v} = (3, 3)$ in $\mathbb{R}^2$, which are shown in Figure 65. We have

$$|\mathbf{u}| = 2$$

and

$$|\mathbf{v}| = \sqrt{3^2 + 3^2} = \sqrt{2 \times 3^2} = 3\sqrt{2}.$$

The angle θ between the vectors $\mathbf{u}$ and $\mathbf{v}$ is $\pi/4$. Hence

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} &= |\mathbf{u}||\mathbf{v}| \cos\theta \\ &= 2 \times 3\sqrt{2} \times \cos\frac{\pi}{4} \\ &= 6\sqrt{2} \times \frac{1}{\sqrt{2}} \\ &= 6.\end{aligned}$$

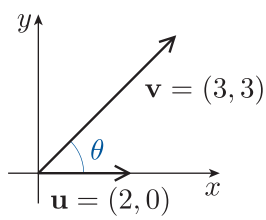


Figure 65 The vectors $\mathbf{u} = (2, 0)$ and $\mathbf{v} = (3, 3)$

There is an easier way to calculate the scalar product of two vectors, which does not depend on knowing the angle between them, but just involves their components. You will meet this method shortly, but first we will use the definition of the scalar product to derive some of its properties.

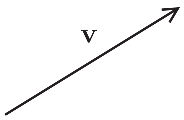


Figure 66 A vector $\mathbf{v}$

To start with, consider any vector $\mathbf{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, as illustrated in Figure 66. Let us find the scalar product of $\mathbf{v}$ with itself. The angle between $\mathbf{v}$ and itself is 0, so we have

$$\mathbf{v} \cdot \mathbf{v} = |\mathbf{v}||\mathbf{v}| \cos 0 = |\mathbf{v}|^2 \times 1 = |\mathbf{v}|^2.$$

This gives the following property.

Magnitude of a vector in terms of scalar product

For any vector $\mathbf{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$,

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

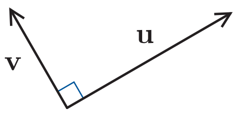


Figure 67 Perpendicular vectors $\mathbf{u}$ and $\mathbf{v}$

Now consider any two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ that are at right angles to each other, as illustrated in Figure 67. Their scalar product is

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos \frac{\pi}{2} = |\mathbf{u}||\mathbf{v}| \times 0 = 0.$$

So the scalar product of any two perpendicular vectors is 0.

A *converse* of this result also holds. Suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ whose scalar product is 0. Then, by the definition of the scalar product,

$$|\mathbf{u}||\mathbf{v}| \cos \theta = 0,$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. It follows that

$$|\mathbf{u}| = 0 \quad \text{or} \quad |\mathbf{v}| = 0 \quad \text{or} \quad \cos \theta = 0,$$

and hence

$$\mathbf{u} = \mathbf{0} \quad \text{or} \quad \mathbf{v} = \mathbf{0} \quad \text{or} \quad \theta = \frac{\pi}{2}.$$

So we have the following property.

Scalar product and perpendicularity

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors.

- If $\mathbf{u}$ and $\mathbf{v}$ are perpendicular, then $\mathbf{u} \cdot \mathbf{v} = 0$.
- If $\mathbf{u} \cdot \mathbf{v} = 0$, then $\mathbf{u} = \mathbf{0}$, or $\mathbf{v} = \mathbf{0}$, or $\mathbf{u}$ and $\mathbf{v}$ are perpendicular.

Finally, the scalar product has the following algebraic properties.

Algebraic properties of scalar product

Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, and let $\alpha \in \mathbb{R}$. The following properties hold.

Commutativity $\quad \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

Multiples $\quad (\alpha\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (\alpha\mathbf{v}) = \alpha(\mathbf{u} \cdot \mathbf{v})$

Distributivity $\quad \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w},$
$\quad (\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}.$

Note that the distributive properties in the box also hold if the plus signs are replaced by minus signs. This follows by combining the distributive properties with the multiples property for $\alpha = -1$.

The properties of the scalar product in the box can be proved by using the definition of the scalar product.

The commutative property follows immediately from the definition.

To see why the multiples property holds, let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, and first suppose that α is a *positive* constant. If the angle between $\mathbf{u}$ and $\mathbf{v}$ is θ, then the angle between $\alpha\mathbf{u}$ and $\mathbf{v}$ is also θ, as illustrated in Figure 68, so

$$\begin{aligned}(\alpha\mathbf{u}) \cdot \mathbf{v} &= |\alpha\mathbf{u}||\mathbf{v}| \cos\theta \\ &= |\alpha||\mathbf{u}||\mathbf{v}| \cos\theta \\ &= \alpha|\mathbf{u}||\mathbf{v}| \cos\theta \qquad \text{(since } \alpha \text{ is positive)} \\ &= \alpha(\mathbf{u} \cdot \mathbf{v}),\end{aligned}$$

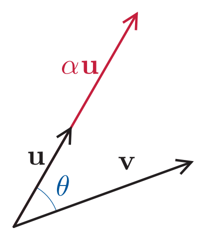


Figure 68 Vectors $\mathbf{u}$ and $\mathbf{v}$, and a scalar multiple $\alpha\mathbf{u}$ of $\mathbf{u}$, where α is positive

and, similarly,

$$\begin{aligned}\mathbf{u} \cdot (\alpha\mathbf{v}) &= |\mathbf{u}||\alpha\mathbf{v}| \cos\theta \\ &= |\mathbf{u}||\alpha||\mathbf{v}| \cos\theta \\ &= \alpha|\mathbf{u}||\mathbf{v}| \cos\theta \\ &= \alpha(\mathbf{u} \cdot \mathbf{v}).\end{aligned}$$

The multiples property can be proved in the case where α is negative in a similar way. In this case the angle between $\alpha\mathbf{u}$ and $\mathbf{v}$, and also the angle between $\mathbf{u}$ and $\alpha\mathbf{v}$, is $\pi - \theta$, but $\cos(\pi - \theta) = -\cos\theta$ by the properties of the cosine function (see the module Handbook).

The proof of the distributive properties is more complicated, and the details are omitted here.

Using the properties of the scalar product given above, we can prove the following simple formulas for calculating the scalar product.

Scalar product of vectors in component form

In $\mathbb{R}^2$, let $\mathbf{u} = (x_1, y_1)$ and $\mathbf{v} = (x_2, y_2)$. Then

$$\mathbf{u} \cdot \mathbf{v} = x_1x_2 + y_1y_2.$$

In $\mathbb{R}^3$, let $\mathbf{u} = (x_1, y_1, z_1)$ and $\mathbf{v} = (x_2, y_2, z_2)$. Then

$$\mathbf{u} \cdot \mathbf{v} = x_1x_2 + y_1y_2 + z_1z_2.$$

Here is a proof of the formula above for vectors in $\mathbb{R}^2$. The proof for vectors in $\mathbb{R}^3$ is similar, but longer.

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^2$. We write them in component form as

$$\mathbf{u} = x_1\mathbf{i} + y_1\mathbf{j} \quad \text{and} \quad \mathbf{v} = x_2\mathbf{i} + y_2\mathbf{j},$$

as shown in Figure 69 below.

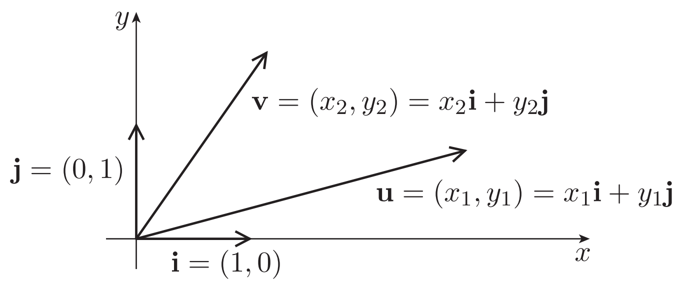


Figure 69 The vectors $\mathbf{u}$ and $\mathbf{v}$ in component form

This gives

$$\begin{aligned}
\mathbf{u} \cdot \mathbf{v} &= (x_1\mathbf{i} + y_1\mathbf{j}) \cdot (x_2\mathbf{i} + y_2\mathbf{j}) \\
&= (x_1\mathbf{i} + y_1\mathbf{j}) \cdot x_2\mathbf{i} + (x_1\mathbf{i} + y_1\mathbf{j}) \cdot y_2\mathbf{j} \quad \text{(by distributivity)} \\
&= x_1\mathbf{i} \cdot x_2\mathbf{i} + y_1\mathbf{j} \cdot x_2\mathbf{i} + x_1\mathbf{i} \cdot y_2\mathbf{j} + y_1\mathbf{j} \cdot y_2\mathbf{j} \\
&\qquad \text{(by distributivity)} \\
&= x_1x_2\mathbf{i} \cdot \mathbf{i} + y_1x_2\mathbf{j} \cdot \mathbf{i} + x_1y_2\mathbf{i} \cdot \mathbf{j} + y_1y_2\mathbf{j} \cdot \mathbf{j} \\
&\qquad \text{(by the multiples property).}
\end{aligned}$$

Now $\mathbf{i}$ and $\mathbf{j}$ have magnitude 1, so by the formula for the magnitude of a vector in terms of scalar product, given earlier,

$$\mathbf{i} \cdot \mathbf{i} = 1^2 = 1 \quad \text{and} \quad \mathbf{j} \cdot \mathbf{j} = 1^2 = 1.$$

Also, $\mathbf{i}$ and $\mathbf{j}$ are perpendicular, so

$$\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{i} = 0.$$

Hence

$$\begin{aligned}
\mathbf{u} \cdot \mathbf{v} &= x_1x_2 \times 1 + y_1x_2 \times 0 + x_1y_2 \times 0 + y_1y_2 \times 1 \\
&= x_1x_2 + y_1y_2,
\end{aligned}$$

as claimed.

Worked Exercise A20

Calculate the following scalar products.

(a) $(3,3)\cdot(2,0)$ (b) $(2\mathbf{i}+3\mathbf{j})\cdot(2\mathbf{i}-\mathbf{j})$ (c) $(\sqrt{2},-4)\cdot(2\sqrt{2},1)$

(d) $(1,-1,1)\cdot(1,-1,1)$

Solution

(a) $$\begin{aligned}(3,3)\cdot(2,0) &= 3\times 2+3\times 0\\ &= 6+0=6\end{aligned}$$

(b) $$\begin{aligned}(2\mathbf{i}+3\mathbf{j})\cdot(2\mathbf{i}-\mathbf{j}) &= 2\times 2+3\times(-1)\\ &= 4-3=1\end{aligned}$$

(c) $$\begin{aligned}(\sqrt{2},-4)\cdot(2\sqrt{2},1) &= \sqrt{2}\times 2\sqrt{2}-4\times 1\\ &= 4-4=0\end{aligned}$$

(d) $$\begin{aligned}(1,-1,1)\cdot(1,-1,1) &= 1\times 1+(-1)\times(-1)+1\times 1\\ &= 1+1+1=3\end{aligned}$$

Worked Exercise A20(a) is the particular scalar product that was calculated using the original definition near the start of this subsection.

Notice that the result of Worked Exercise A20(c) shows that the vectors $(\sqrt{2},-4)$ and $(2\sqrt{2},1)$ are perpendicular, something that is not immediately obvious when we look at their component forms.

Exercise A51

Calculate the following scalar products.

(a) $(2,3)\cdot\left(\frac{5}{2},-4\right)$ (b) $(1,4)\cdot\left(2,-\frac{1}{2}\right)$ (c) $(-2\mathbf{i}+\mathbf{j})\cdot(3\mathbf{i}-2\mathbf{j})$

(d) $(1,-1,-2)\cdot(3,-2,-5)$

One useful application of the scalar product is that it provides a method for finding the angle between two vectors, as illustrated in Figure 70. The formula below is obtained by rearranging the original definition of the scalar product.

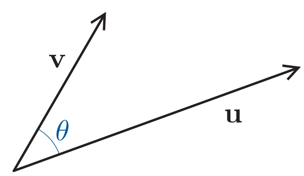


Figure 70 Two vectors $\mathbf{u}$ and $\mathbf{v}$, and the angle θ between them

Angle between two vectors

The angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\cos\theta=\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|}.$$

In the next worked exercise this formula is used to find the angle between two vectors in $\mathbb{R}^2$.

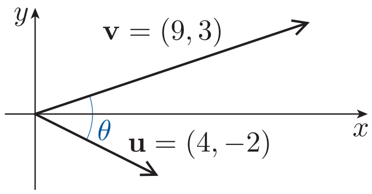


Figure 71 The vectors $\mathbf{u} = (4,-2)$ and $\mathbf{v} = (9,3)$

Worked Exercise A21

Find the angle θ between the vectors $\mathbf{u} = (4,-2)$ and $\mathbf{v} = (9,3)$, in radians. (These vectors are shown in Figure 71.)

Solution

We have

$$\mathbf{u} \cdot \mathbf{v} = 4 \times 9 + (-2) \times 3 = 36 - 6 = 30,$$

$$|\mathbf{u}| = \sqrt{4^2 + (-2)^2} = \sqrt{20} = 2\sqrt{5},$$

$$|\mathbf{v}| = \sqrt{9^2 + 3^2} = \sqrt{90} = 3\sqrt{10}.$$

So

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \frac{30}{2\sqrt{5} \times 3\sqrt{10}} = \frac{30}{30\sqrt{2}} = \frac{1}{\sqrt{2}}.$$

For the angle between the two vectors, we need to choose the value of θ that satisfies this equation and lies in the range $0 \leq \theta \leq \pi$.

Hence

$$\theta = \cos^{-1}\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{4}.$$

Exercise A52

Find the angle between the vectors in each of the following pairs of vectors, in radians. Give your answer to two decimal places unless it is an obvious multiple of π.

(a) $(1,4)$, $(5,2)$ (b) $(-2,2)$, $(1,-1)$ (c) $9\mathbf{i} - 2\mathbf{j}$, $\mathbf{i} + 2\mathbf{j}$

The formula for the angle between two vectors works equally well in $\mathbb{R}^3$, as is shown in the next worked exercise.

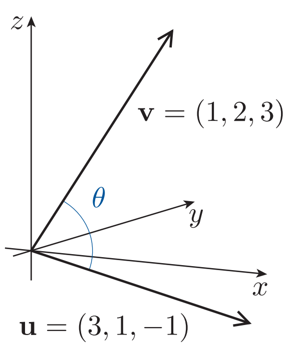


Figure 72 The vectors $\mathbf{u} = (3,1,-1)$ and $\mathbf{v} = (1,2,3)$

Worked Exercise A22

Find the angle θ between the vectors $\mathbf{u} = (3,1,-1)$ and $\mathbf{v} = (1,2,3)$, in radians to two decimal places. (These vectors are shown in Figure 72.)

Solution

We have

$$\mathbf{u} \cdot \mathbf{v} = 3 \times 1 + 1 \times 2 + (-1) \times 3 = 2,$$

$$|\mathbf{u}| = \sqrt{3^2 + 1^2 + (-1)^2} = \sqrt{11},$$

$$|\mathbf{v}| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}.$$

So

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \frac{2}{\sqrt{11}\sqrt{14}} = \frac{2}{\sqrt{154}}.$$

Hence

$$\theta = \cos^{-1}\left(\frac{2}{\sqrt{154}}\right) = 1.41 \text{ radians (to 2 d.p.)}.$$

Exercise A53

Find the angle between the following pairs of vectors, in radians to two decimal places.

(a) $(3, 4, 5)$, $(1, 0, -1)$

(b) $2\mathbf{j} - 3\mathbf{k}$, $-\mathbf{i} - \mathbf{j} - 2\mathbf{k}$

4.5 Equation of a plane in $\mathbb{R}^3$

In Subsection 1.1 you saw that the general form of the equation of a line in $\mathbb{R}^2$ is $ax + by = c$, where $a, b, c \in \mathbb{R}$, and a and b are not both zero. We can use the scalar product to derive a similar general form for the equation of a plane in $\mathbb{R}^3$, as you will see in this subsection. In doing this, we will also derive a general form for the equation of a plane in $\mathbb{R}^3$ in terms of vectors.

First, let us look at some planes in $\mathbb{R}^3$ whose equations are easy to find. The 'simplest' planes in $\mathbb{R}^3$ are the three planes that contain a pair of axes. The (x, y)**-plane** is the plane that contains the x- and y-axes, as illustrated in Figure 73. The (x, z)**-plane** and the (y, z)**-plane** are defined similarly. The points that lie in the (x, y)-plane are the points (x, y, z) in $\mathbb{R}^3$ for which $z = 0$, so the equation of the (x, y)-plane is

$$z = 0.$$

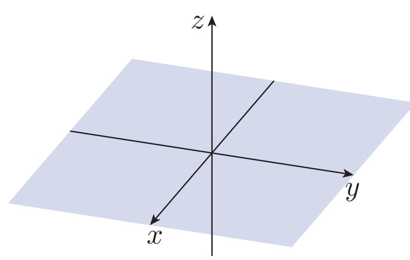


Figure 73 The (x, y)-plane

Exercise A54

Write down the equations of the (y, z)-plane and the (x, z)-plane.

Exercise A55

Sketch the planes whose equations are as follows.

(a) $z = 2$ (b) $y = -1$

Before we derive the general equation of a plane in $\mathbb{R}^3$, we need the following concept.

Definition

A vector that is perpendicular to all the vectors in a particular plane is called a **normal vector** (or simply a **normal**) to the plane. Its direction is said to be **normal** to the plane.

Figure 74(a) shows some normal vectors to a plane. If $\mathbf{n}$ is a normal vector to a particular plane, then so is $k\mathbf{n}$, for any non-zero real number k. If $k > 0$, then $k\mathbf{n}$ is in the same direction as $\mathbf{n}$, whereas if $k < 0$, then $k\mathbf{n}$ is in the opposite direction to $\mathbf{n}$, as illustrated in Figure 74(b).

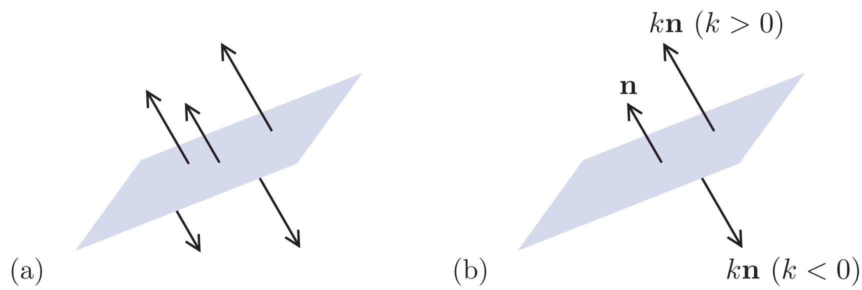


Figure 74 Some normal vectors to a plane

Any vector $\mathbf{n}$ in $\mathbb{R}^3$ is a normal vector to infinitely many planes, all parallel to each other, as illustrated in Figure 75.

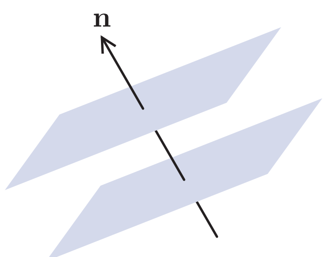


Figure 75 Parallel planes

We can specify any particular plane in $\mathbb{R}^3$ by specifying a normal vector to the plane, together with a point that lies in the plane. For example, there is exactly one plane that contains the point $P(2, 3, 4)$ and has $\mathbf{n} = (1, 2, -1)$ as a normal.

Here is how we can find an equation for this particular plane. A condition for an arbitrary point $X(x, y, z)$ in $\mathbb{R}^3$ to lie in the plane is that the vector $\overrightarrow{PX}$ must be perpendicular to the normal vector $\mathbf{n}$, as illustrated in Figure 76. In other words, we must have

$$\overrightarrow{PX} \cdot \mathbf{n} = 0.$$

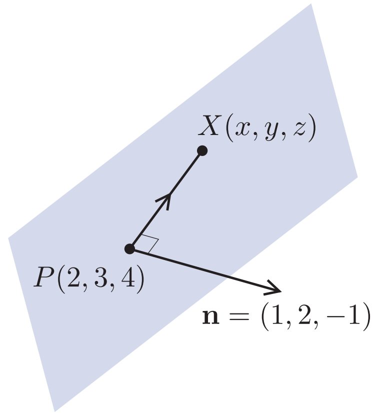


Figure 76 An arbitrary point $X(x, y, z)$ on the plane containing the point $P(2, 3, 4)$ with normal $\mathbf{n} = (1, 2, -1)$

Now

$$\begin{aligned}\overrightarrow{PX} &= \mathbf{x} - \mathbf{p} \quad \text{(where } \mathbf{x} \text{ and } \mathbf{p} \text{ are the position vectors of } X \text{ and } P\text{)}\\ &= (x, y, z) - (2, 3, 4)\\ &= (x - 2, y - 3, z - 4).\end{aligned}$$

Hence the condition for the point $X(x, y, z)$ to lie in the plane is

$$(x - 2, y - 3, z - 4) \cdot (1, 2, -1) = 0,$$

that is,

$$(x - 2) \times 1 + (y - 3) \times 2 + (z - 4) \times (-1) = 0,$$

which simplifies to

$$x + 2y - z = 4.$$

This is the equation of the plane.

In fact every plane in $\mathbb{R}^3$ has an equation of the form

$$ax + by + cz = d,$$

for some real numbers a, b, c and d. To prove this, we apply the argument above to a general plane. Consider the plane that contains the point $P(x_1, y_1, z_1)$ and has $\mathbf{n} = (a, b, c)$ as a normal vector, as illustrated in Figure 77.

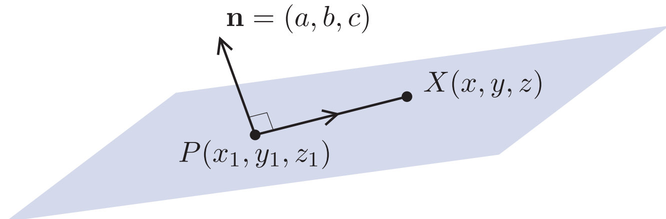


Figure 77 An arbitrary point $X(x, y, z)$ on the plane containing the point $P(x_1, y_1, z_1)$ with normal $\mathbf{n} = (a, b, c)$

A condition for an arbitrary point $X(x,y,z)$ in $\mathbb{R}^3$ to lie in this plane is that the vectors $\overrightarrow{PX}$ and $\mathbf{n}$ must be perpendicular, that is,

$$\overrightarrow{PX} \cdot \mathbf{n} = 0.$$

Since $\overrightarrow{PX} = \mathbf{x} - \mathbf{p}$, where $\mathbf{x}$ and $\mathbf{p}$ are the position vectors of X and P, respectively, this condition can be written as

$$(\mathbf{x} - \mathbf{p}) \cdot \mathbf{n} = 0.$$

By the algebraic properties of the scalar product, we can write the condition as

$$\mathbf{x} \cdot \mathbf{n} - \mathbf{p} \cdot \mathbf{n} = 0,$$

that is,

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{p} \cdot \mathbf{n}.$$

This is the vector form of the equation of the plane. Alternatively, we can write it in terms of the coordinates x, y and z, by substituting for $\mathbf{x} = (x,y,z)$, $\mathbf{n} = (a,b,c)$ and $\mathbf{p} = (x_1,y_1,z_1)$. Then the equation becomes

$$(x,y,z) \cdot (a,b,c) = (x_1,y_1,z_1) \cdot (a,b,c),$$

that is,

$$ax + by + cz = ax_1 + by_1 + cz_1.$$

This equation is of the form

$$ax + by + cz = d,$$

where d is the real number given by $d = ax_1 + by_1 + cz_1$. So we have shown that every plane in $\mathbb{R}^3$ has an equation of this form, for some real numbers a, b, c and d.

Equation of a plane in $\mathbb{R}^3$

The equation of the plane that contains the point (x_1,y_1,z_1) and has the vector $\mathbf{n} = (a,b,c)$ as a normal is

$$ax + by + cz = d,$$

where $d = ax_1 + by_1 + cz_1$.

This equation can be written in vector form as

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{p} \cdot \mathbf{n},$$

where $\mathbf{x} = (x,y,z)$ and $\mathbf{p} = (x_1,y_1,z_1)$.

Once we know the equation of a plane in the form $ax + by + cz = d$, we can 'read off' the components of a normal vector, as they are the coefficients of x, y and z in the equation. For instance, one normal to the plane with equation $x - 2y + 3z = 7$ is $\mathbf{n} = (1,-2,3)$. Note that the zero vector can never be a normal since its direction is undefined.

When we want to find the equation of a plane, it is simpler to start from the vector form of the equation, as demonstrated in the next worked exercise.

Worked Exercise A23

Determine the equation of the plane in $\mathbb{R}^3$ that contains the point $(1, -1, 4)$ and has the vector $(2, -2, 3)$ as a normal.

Solution

The equation of the plane is

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{p} \cdot \mathbf{n},$$

where $\mathbf{x} = (x, y, z)$, $\mathbf{p} = (1, -1, 4)$ and $\mathbf{n} = (2, -2, 3)$. So the equation of the plane is

$$(x, y, z) \cdot (2, -2, 3) = (1, -1, 4) \cdot (2, -2, 3),$$

that is,

$$2x - 2y + 3z = 1 \times 2 + (-1) \times (-2) + 4 \times 3,$$

which simplifies to

$$2x - 2y + 3z = 16.$$

Exercise A56

Determine the equation of each of the following planes.

(a) The plane that contains the point $(1, 0, 2)$ and has the vector $(2, 3, 1)$ as a normal.

(b) The plane that contains the point $(-1, 1, 5)$ and has the vector $(4, -2, 1)$ as a normal.

In Book C you will see how you can find the equation of a plane in $\mathbb{R}^3$ if you know three points on the plane, rather than a point and a normal.

Summary

In this unit you have studied some fundamental ideas in mathematics. You have met a new notation for specifying sets and encountered examples of sets of numbers and sets of points. You have studied the operations of union, intersection and difference that can be performed on sets, and seen how to show that two sets are equal. You have also met many examples of functions between sets, and seen that a one-to-one function has an inverse. Finally, you have worked with vectors and seen how to carry out vector arithmetic in component form and use the scalar product of two vectors.

Throughout the unit you have worked especially with the sets $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$, of real numbers, ordered pairs of real numbers and ordered triples of real numbers, respectively. You have seen that the elements of these sets can be regarded geometrically as points on the real line, in the plane and in space, and that points in $\mathbb{R}^2$ or $\mathbb{R}^3$ can also be identified with their position vectors.

You will continue your study of foundational mathematical concepts in the rest of Book A, and the ideas you meet here will be in constant use throughout this module.

Learning outcomes

After working through this unit, you should be able to:

- recognise the *equation of a line* and the *equation of a circle* in $\mathbb{R}^2$
- use *set notation* and the notation of *intervals* of the real line
- determine whether one set is a subset of another, and whether two sets are equal
- find the *union*, *intersection* and *difference* of two sets
- define a *function* and its *domain*, *codomain* and *rule*
- determine the *image set* of a function
- determine whether a function is *one-to-one* and/or *onto*
- find the *inverse* of a one-to-one function, and the *composite* of two functions
- explain what are meant by a *vector*, a *scalar*, a *scalar multiple* of a vector, and the *sum* and *difference* of two vectors
- represent vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ in terms of their *components*, and carry out vector arithmetic using components
- determine the equation of a line in $\mathbb{R}^2$ or $\mathbb{R}^3$ in terms of vectors
- explain what is meant by the *scalar product* of two vectors, and use it to find the angle between two vectors
- recognise the *equation of a plane* in $\mathbb{R}^3$, and the vector form of the equation
- determine the equation of a plane in $\mathbb{R}^3$, given a point in the plane and a *normal* to the plane.

Solutions to exercises

Solution to Exercise A1

Using the formula for the equation of a line when given its gradient and one point on it, we find that the equation of this line is

$$y - (-1) = -3(x - 2).$$

We can rearrange this to

$$y = -3x + 5,$$

or

$$3x + y = 5.$$

Solution to Exercise A2

(a) Since $(1, 1)$ and $(3, 5)$ lie on the line, its gradient is

$$m = \frac{1-5}{1-3} = 2.$$

Then, since the point $(1, 1)$ lies on the line, its equation must be

$$y - 1 = 2(x - 1),$$

so

$$y = 2x - 1, \text{ or } \quad 2x - y = 1.$$

(b) Both these points have x-coordinate 0, so they lie on the line with equation $x = 0$, the y-axis.

(c) Since the origin lies on the line, its equation must be of the form $y = mx$, where m is its gradient.

Since $(4, 2)$ lies on the line, its coordinates must satisfy the equation of the line. Thus $2 = 4m$, so $m = \frac{1}{2}$.

Hence the equation of this line is $y = \frac{1}{2}x$, or $\frac{1}{2}x - y = 0$, or $x - 2y = 0$.

(d) Both these points have y-coordinate -1, so they lie on the line with equation $y = -1$.

Solution to Exercise A3

We can rearrange the equations of the lines to find their gradients as follows:

$$\begin{array}{ll} l_1: \ y = -2x + 4 & l_2: \ y = 2x + \frac{4}{3} \\ l_3: \ y = -\frac{1}{2}x + 5 & l_4: \ y = \frac{1}{2}x - \frac{5}{6} \\ l_5: \ y = \frac{1}{2}x + 1 & l_6: \ y = -2x - \frac{7}{2} \end{array}$$

Thus the gradients of the given lines are -2, 2, $-\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$ and -2, respectively.

It follows that the lines l_1 and l_6 are parallel, since their gradients are the same but their y-intercepts are different. Similarly, l_4 and l_5 are also parallel.

Lines l_1 and l_4 are perpendicular, since the product of their gradients is -1. For the same reason, each of the following pairs of lines are perpendicular: l_1 and l_5; l_2 and l_3; l_4 and l_6; and l_5 and l_6.

Solution to Exercise A4

We use the formula for the distance between two points in the plane. This gives the following distances.

(a) $\sqrt{(5-0)^2 + (0-0)^2} = 5$

(b) $\sqrt{(3-0)^2 + (4-0)^2} = 5$

(c) $\sqrt{(5-1)^2 + (1-2)^2} = \sqrt{17}$

(d) $\sqrt{(-1-3)^2 + (4-(-8))^2} = \sqrt{160}$
$= 4\sqrt{10}$

(The two points in part (a) are on the x-axis, so in fact there is no need to use the distance formula to find the distance between them.)

Solution to Exercise A5

(a) This circle has equation

$$(x - 0)^2 + (y - 0)^2 = 4^2,$$

which can be simplified to give

$$x^2 + y^2 = 16.$$

(b) This circle has equation

$$(x - (-1))^2 + (y - 0)^2 = (\sqrt{2})^2,$$

which can be simplified to give

$$(x + 1)^2 + y^2 = 2.$$

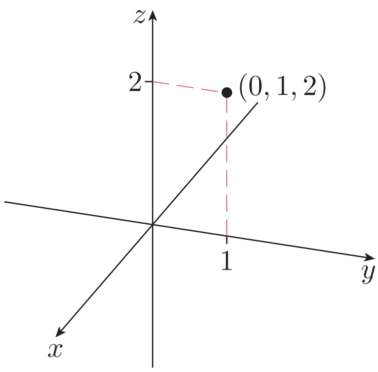
(c) This circle has equation

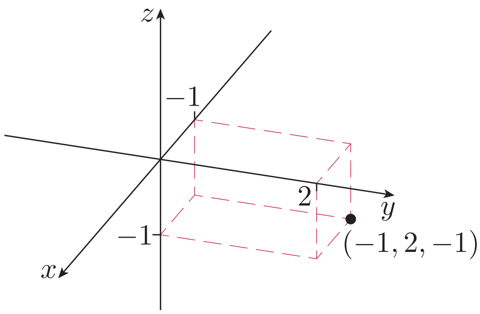
$$(x - 3)^2 + (y - (-4))^2 = 2^2,$$

which can be simplified to give

$$(x - 3)^2 + (y + 4)^2 = 4.$$

Solution to Exercise A6





Solution to Exercise A7

We use the formula for the distance between two points in $\mathbb{R}^3$. This gives the following distances.

(a) $\sqrt{(4-1)^2+(1-1)^2+(-3-1)^2}$
$= \sqrt{9+0+16} = 5$

(b) $\sqrt{(3-1)^2+(0-2)^2+(3-3)^2}$
$= \sqrt{4+4+0} = 2\sqrt{2}$

Solution to Exercise A8

(a) True: -3 is an integer.

(b) False: 5 is a natural number.

(c) False: 1.3 is the rational number $\frac{13}{10}$.

(d) True: both 1 and 3 are rational numbers.

(e) True: $-\pi$ is a real number.

(f) False: $\frac{1}{2}$ is not a natural number.

(g) False: 1 is a non-zero real number, but 0 is not.

(h) False: $\sqrt{2}$ is a real number.

Solution to Exercise A9

(a) True: 1 is a member of the given set.

(b) True: the set $\{-9\}$ is a member of the given set, although the number -9 is not.

(c) False: the number 9 belongs to the given set, but the set $\{9\}$ does not.

(d) False: the point $(0, 1)$ is not a member of the given set of points in $\mathbb{R}^2$, although the point $(1, 0)$ is.

(e) False: the numbers 1 and 0 are not members of the given set of points in $\mathbb{R}^2$, although the point $(1, 0)$ is.

(f) True: the set $\{1, 0\}$ is the same as the set $\{0, 1\}$, and so is a member of the given set. Notice that the members of this set are themselves *sets*, and not points in $\mathbb{R}^2$.

Solution to Exercise A10

(a) True: $\frac{9}{2}$ is in $\mathbb{R}$, and it satisfies the condition $x > 3$.

(b) True: $7 = 3 \times 2 + 1$, so 7 is of the form $3k + 1$ for some $k \in \mathbb{Z}$.

(c) False: $-\frac{7}{2}$ is not in $\mathbb{Z}$.

(d) False: 8 cannot be expressed as 2^x for some number $x \in \mathbb{R}$ satisfying $0 < x < 2$; in fact $8 = 2^3$.

(e) True: 9 is in $\mathbb{Z}$, and $9 = 3^2$ so $9 = k^2$ for some $k \in \mathbb{Z}$.

(f) True: $6 = 3(3-1)$, so 6 is of the form $m(m-1)$ for some $m \in \mathbb{N}$.

(g) False: 4 is an even integer, but it does not satisfy $0 < r < 4$.

Solution to Exercise A11

(a) $\{k \in \mathbb{Z} : -2 < k < 1000\}$

(b) $\{x \in \mathbb{Q} : x > 0 \text{ and } x^2 > 2\}$ or $\{x \in \mathbb{Q} : x > 0,\ x^2 > 2\}$

(c) $\{2n : n \in \mathbb{N}\}$

(d) $\{2^k : k \in \mathbb{Z}\}$

Solution to Exercise A12

(a) False: the set $(1, 5)$ is an open interval and does not include the endpoint 1.

(b) True: the set $(-1, 1]$ is half-open, with the upper endpoint 1 included.

(c) False: ∞ does not denote a number and so is not in the interval.

(d) True: $\mathbb{R}^*$ denotes the set of non-zero real numbers, so 0 is not a member of this set.

(e) False: $x \in \mathbb{R}^*$ means x is a non-zero real number, while $(0, \infty)$ comprises just the positive real numbers. For example, the number -1 is in $\mathbb{R}^*$, but not in $(0, \infty)$.

Solution to Exercise A13

(a) $[-11, 2)$

(b) $(-6.5, 21]$

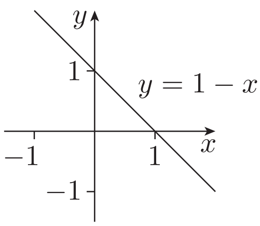
(c) $(-273, \infty)$

Solution to Exercise A14

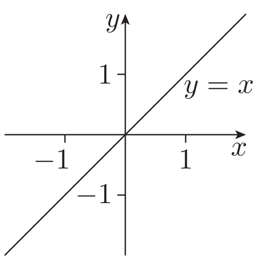
(a) $l = \{(x, y) \in \mathbb{R}^2 : y = 2x + 5\}$

(There are other ways to specify this line; another example is $l = \{(x, 2x + 5) : x \in \mathbb{R}\}$.)

(b) The line l has equation $y = 1 - x$, so it is as follows.



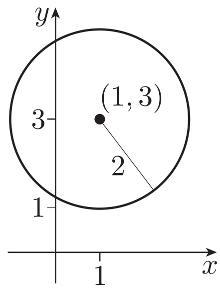
(c) The line l has equation $y = x$ (since here $m = 1$ and $c = 0$), so it is as follows.



Solution to Exercise A15

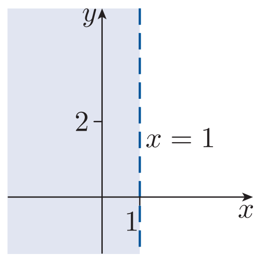
(a) $C = \{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + (y + 4)^2 = 9\}$

(b) The circle has centre $(1, 3)$ and radius 2.

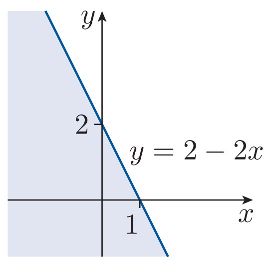


Solution to Exercise A16

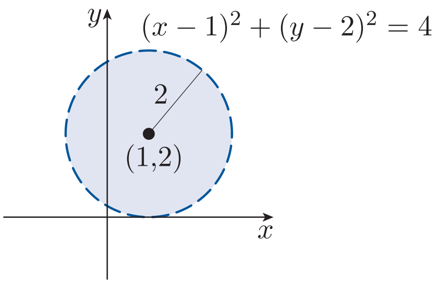
(a) This set is a half-plane with the boundary line excluded, as follows.



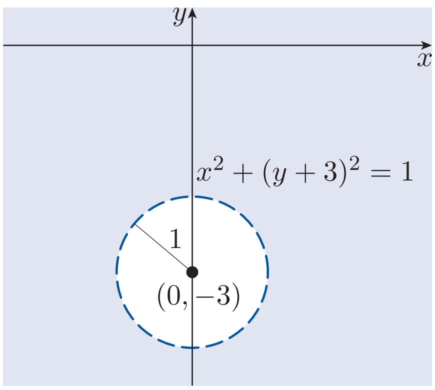
(b) This set is another half-plane, but this time the boundary line is included, as follows.



(c) This set is a disc with the boundary excluded, as follows.

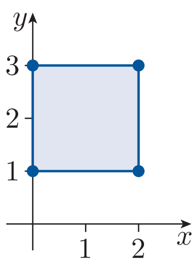


(d) This set consists of the points outside a disc with centre $(0,-3)$ and radius 1, as follows.



Solution to Exercise A17

$\{(x,y)\in\mathbb{R}^2 : 0\leq x\leq 2,\ 1\leq y\leq 3\}$



Solution to Exercise A18

(a) The set B consists of the solutions of the equation

$$x^2+x-6=0,$$

which we can write as

$$(x-2)(x+3)=0.$$

So $B=\{2,-3\}=A$.

(b) The two sets are

$$\begin{aligned}A&=\{k\in\mathbb{Z} : k \text{ is odd and } 0<k<8\}\\&=\{1,3,5,7\},\\B&=\{2n+1 : n\in\mathbb{N} \text{ and } n^2<25\}\\&=\{3,5,7,9\}.\end{aligned}$$

Hence $A\neq B$, either because $9\in B$ but $9\notin A$, or because $1\in A$ but $1\notin B$.

Solution to Exercise A19

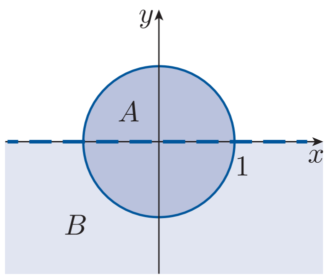
(a) Each element of A is a point in $\mathbb{R}^2$.

We calculate $x-4y$ using the coordinates of each point of A:

$$\begin{aligned}5-4\times 2&=-3,\\1-4\times 1&=-3,\\-3-4\times 0&=-3.\end{aligned}$$

This shows that each element of A is an element of B, so $A\subseteq B$.

(b) The sets A and B are sketched below.



The set A is the interior of the unit circle, and B is the half-plane consisting of all points with negative y-coordinate. So $A\not\subseteq B$, because, for example, the point $\left(\frac{1}{2},\frac{1}{2}\right)$ belongs to A but not to B. (Any one point that is in set A but not in set B shows that $A\not\subseteq B$.)

(c) Let x be an arbitrary element of A; then $x\in\mathbb{R}$ and satisfies $-1\leq x\leq 0$. This equation gives

$$-1+1\leq x+1\leq 0+1,$$

that is,

$$0\leq x+1\leq 1.$$

Hence

$$0\leq (x+1)^2\leq 1,$$

so $x\in B$.

Since x is an arbitrary element of A, we conclude that $A\subseteq B$.

Solution to Exercise A20

(a) We showed that $A \subseteq B$ in the solution to Exercise A19(a). Also, for example, the point $(9,3)$ lies in B, since

$$9 - 4 \times 3 = -3,$$

but does not lie in A. Therefore A is a proper subset of B.

(b) We showed that $A \subseteq B$ in the solution to Exercise A19(c). Also, for example, -2 lies in B, since

$$(-2+1)^2 = (-1)^2 = 1,$$

but does not lie in A. Therefore A is a proper subset of B.

Solution to Exercise A21

(a) First we show that $A \subseteq B$.

Let $(x,y) \in A$; then $(x,y) \in \mathbb{R}^2$, and for some $t \in \mathbb{R}$, we have $x = t^2$ and $y = 2t$. Hence

$$y^2 = (2t)^2 = 4t^2 = 4x.$$

So $(x,y) \in B$, and $A \subseteq B$.

Next we show that $B \subseteq A$.

Let $(x,y) \in B$; then $y^2 = 4x$. We must show that there is a value of t in $\mathbb{R}$ such that $x = t^2$ and $y = 2t$, so that $(x,y) \in A$. Let t be given by $y = 2t$; that is, $t = \frac{1}{2}y$. Then, since $4x = y^2$, we have $x = \frac{1}{4}y^2$, and substituting for y gives

$$x = \frac{1}{4}(2t)^2 = t^2.$$

Hence $(x,y) = (t^2, 2t) \in A$, and so $B \subseteq A$.

Since $A \subseteq B$ and $B \subseteq A$, it follows that $A = B$.

(b) First we show that $A \subseteq B$.

Let $(x,y) \in A$; then $2x + y - 3 = 0$. We must show that there is a value of t in $\mathbb{R}$ such that $x = t+1$ and $y = 1-2t$. Let t be given by $x = t+1$, that is, $t = x - 1$. Then, since $2x + y - 3 = 0$, we have

$$\begin{aligned} y &= 3 - 2x \\ &= 3 - 2(t+1) \\ &= 1 - 2t. \end{aligned}$$

Hence $(x,y) = (t+1, 1-2t) \in B$, and so $A \subseteq B$.

Next we show that $B \subseteq A$.

Let $(x,y) \in B$; then $(x,y) \in \mathbb{R}^2$, and for some $t \in \mathbb{R}$, we have $x = t+1$ and $y = 1-2t$. We must show that (x,y) satisfies $2x + y - 3 = 0$. Now

$$\begin{aligned} 2x + y - 3 &= 2(t+1) + (1-2t) - 3 \\ &= 0, \end{aligned}$$

as required, so $(x,y) \in A$. Therefore $B \subseteq A$.

Since $A \subseteq B$ and $B \subseteq A$, it follows that $A = B$.

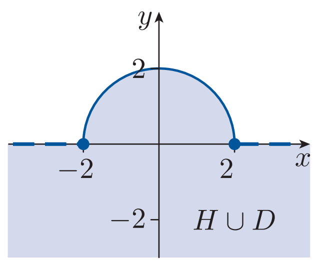
Solution to Exercise A22

(a) $(1,7) \cup [4,11] = (1,11]$.

(b) $\mathbb{R}^*$ denotes the set of non-zero real numbers, and so is the union of the two intervals $(-\infty, 0)$ and $(0, \infty)$; that is

$$\mathbb{R}^* = (-\infty, 0) \cup (0, \infty).$$

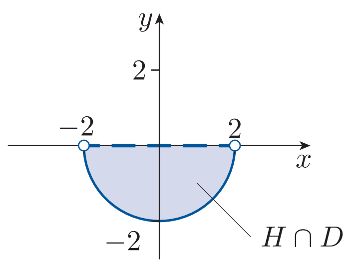
(c) The union of the half-plane and disc is



Solution to Exercise A23

(a) $(1,7) \cap [4,11] = [4,7)$.

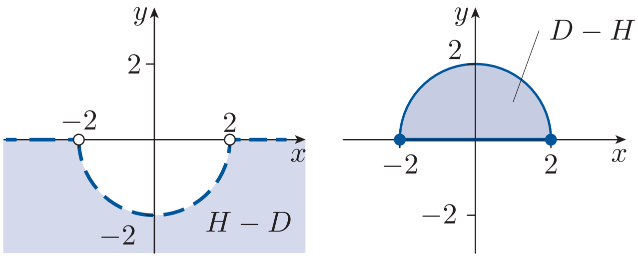
(b) The intersection is



Solution to Exercise A24

(a) $(1,7) - [4,11] = (1,4)$ and $[4,11] - (1,7) = [7,11]$.

(b) The two differences are



Solution to Exercise A25

(a) This is the translation of the plane that moves each point to the right by 2 units and up by 3 units.

(b) This is the reflection of the plane in the x-axis.

(c) This is the rotation of the plane through $\pi/2$ anticlockwise about the origin.

Solution to Exercise A26

Only diagram (b) represents a function.

Diagram (a) does not represent a function, as there is no arrow from the element 3.

Diagram (c) does not represent a function, as there are two arrows from the element 1.

Solution to Exercise A27

(a) $f(S) = \{f(0), f(1), f(2), f(3)\}$
$= \{-1, 0, 1, 2\}$.

(b) $f(\mathbb{Z}) = \{\ldots, f(-1), f(0), f(1), f(2), \ldots\}$
$= \{\ldots, -2, -1, 0, 1 \ldots\}$
$= \mathbb{Z}$.

Solution to Exercise A28

The images of the elements of A are
$f(0) = 9,\ f(1) = 8,\ f(2) = 7,\ f(3) = 6,$
$f(4) = 5,\ f(5) = 4,\ f(6) = 3,\ f(7) = 2,$
$f(8) = 1,\ f(9) = 0.$

So the image set of f is
$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} = A$.

Solution to Exercise A29

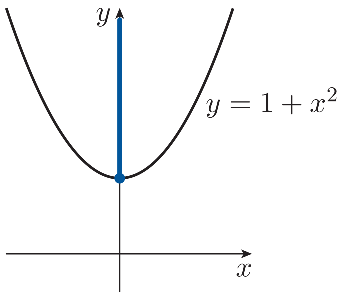
Only diagram (a) represents an onto function.

Diagram (b) does not even represent a function, as there is no arrow from the element 4.

Diagram (c) represents a function that is not onto, as there is no arrow going to the element 1.

Solution to Exercise A30

(a) The sketch of the graph of f below suggests that $f(\mathbb{R}) = [1, \infty)$.



We prove that $f(\mathbb{R}) = [1, \infty)$.

Let $x \in \mathbb{R}$; then $f(x) = 1 + x^2$. Since $x^2 \geq 0$, we have $1 + x^2 \geq 1$ and so $f(\mathbb{R}) \subseteq [1, \infty)$.

We must show that $f(\mathbb{R}) \supseteq [1, \infty)$.

Let $y \in [1, \infty)$. We must show that there exists $x \in \mathbb{R}$ such that $f(x) = y$; that is, $1 + x^2 = y$. Now $x = \sqrt{y-1}$ is real, since $y \geq 1$, and satisfies $f(x) = y$, as required. (Alternatively, $x = -\sqrt{y-1}$ is real and satisfies $f(x) = y$.)

Thus $f(\mathbb{R}) \supseteq [1, \infty)$.

Since $f(\mathbb{R}) \subseteq [1, \infty)$ and $f(\mathbb{R}) \supseteq [1, \infty)$, it follows that $f(\mathbb{R}) = [1, \infty)$, so the image set of f is $[1, \infty)$, as expected.

The interval $[1, \infty)$ is not the whole of the codomain $\mathbb{R}$, so f is not onto.

(b) This function is the reflection of the plane in the x-axis. This suggests that $f(\mathbb{R}^2) = \mathbb{R}^2$.

We know that $f(\mathbb{R}^2) \subseteq \mathbb{R}^2$, so we must show that $f(\mathbb{R}^2) \supseteq \mathbb{R}^2$.

Let $(x', y') \in \mathbb{R}^2$. We must show that there exists $(x, y) \in \mathbb{R}^2$ such that $f(x, y) = (x', y')$; so $(x, -y) = (x', y')$, that is,

$$x = x', \quad -y = y'.$$

Rearranging these equations, we obtain

$$x = x', \quad y = -y'.$$

So, $(x, y) \in \mathbb{R}^2$ and $f(x, y) = (x', y')$, as required.

Thus $f(\mathbb{R}^2) \supseteq \mathbb{R}^2$.

Since $f(\mathbb{R}^2) \subseteq \mathbb{R}^2$ and $f(\mathbb{R}^2) \supseteq \mathbb{R}^2$, it follows that $f(\mathbb{R}^2) = \mathbb{R}^2$, so the image set of f is $\mathbb{R}^2$, as expected.

The codomain of f is also $\mathbb{R}^2$, so f is onto.

Solution to Exercise A31

Only diagram (c) represents a one-to-one function.

Diagram (a) represents a function that is not one-to-one, as there are two arrows going to the element 3.

Diagram (b) does not even represent a function, as there is no arrow from the element 2.

Solution to Exercise A32

(a) This function is not one-to-one since, for example,

$$f(2) = f(-2) = 1 + 4 = 5.$$

(b) This function is the reflection of the plane in the x-axis, so we expect it to be one-to-one. We now prove this algebraically.

Suppose that $f(x_1, y_1) = f(x_2, y_2)$; then

$$(x_1, -y_1) = (x_2, -y_2).$$

This means that $x_1 = x_2$ and $-y_1 = -y_2$,. It follows that $y_1 = y_2$, so we have shown that $(x_1, y_1) = (x_2, y_2)$, that is, f is one-to-one.

Solution to Exercise A33

(a) In Exercise A32 we saw that f is not one-to-one, so f does not have an inverse function.

(b) In Exercise A32 we saw that f is one-to-one, so f has an inverse function.

In Exercise A30 we saw that the image set of f is $\mathbb{R}^2$ and, for each $(x', y') \in \mathbb{R}^2$, we have

$$(x', y') = f(x', -y').$$

So f^{-1} is the function

$$\begin{aligned} f^{-1} : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x', y') &\longmapsto (x', -y'). \end{aligned}$$

This can be expressed in terms of x and y as

$$\begin{aligned} f^{-1} : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x, -y). \end{aligned}$$

(In this case, f^{-1} is actually equal to f, which is what we would expect for a reflection.)

(c) The graph of this function is an upward sloping straight line, which suggests that it is one-to-one. First we confirm this algebraically. Suppose that $f(x_1) = f(x_2)$; then

$$8x_1 + 3 = 8x_2 + 3,$$

so $8x_1 = 8x_2$, and hence $x_1 = x_2$. Thus f is one-to-one, and so it has an inverse function.

We now find the image set of f. We suspect that its image set is $\mathbb{R}$, so we now prove this algebraically. Let y be an arbitrary element in $\mathbb{R}$. We must show that there exists an element x in the domain $\mathbb{R}$ such that

$$f(x) = y; \quad \text{that is,} \quad 8x + 3 = y.$$

Rearranging this equation, we obtain

$$x = \frac{y - 3}{8}.$$

This is in $\mathbb{R}$ and satisfies $f(x) = y$, as required. Thus the image set of f is $\mathbb{R}$.

Hence f^{-1} is the function

$$\begin{aligned} f^{-1} : \mathbb{R} &\longrightarrow \mathbb{R} \\ y &\longmapsto \frac{y - 3}{8}. \end{aligned}$$

This can be expressed in terms of x as

$$\begin{aligned} f^{-1} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{x - 3}{8}. \end{aligned}$$

Solution to Exercise A34

The function

$$\begin{aligned} g : [-\pi/2, \pi/2] &\longrightarrow [-1, 1] \\ x &\longmapsto \sin x \end{aligned}$$

is a restriction of f that is one-to-one.

(There are many other possibilities, for example, the restriction of the domain to $[\pi/2, 3\pi/2]$.)

Solution to Exercise A35

(a) The rule of $g \circ f$ is

$$\begin{aligned}(g \circ f)(x) = g(f(x)) &= g(-x)\\ &= 3(-x) + 1\\ &= -3x + 1.\end{aligned}$$

Thus $g \circ f$ is the function

$$\begin{aligned}g \circ f : \mathbb{R} &\longrightarrow \mathbb{R}\\ x &\longmapsto -3x + 1.\end{aligned}$$

(b) The rule of $f \circ g$ is

$$\begin{aligned}(f \circ g)(x) = f(g(x)) &= f(3x + 1)\\ &= -(3x + 1)\\ &= -3x - 1.\end{aligned}$$

Thus $f \circ g$ is the function

$$\begin{aligned}f \circ g : \mathbb{R} &\longrightarrow \mathbb{R}\\ x &\longmapsto -3x - 1.\end{aligned}$$

Solution to Exercise A36

The rule of $f \circ g$ is

$$\begin{aligned}(f \circ g)(x, y) = f(g(x, y)) &= f(-x, y)\\ &= (-x, -y).\end{aligned}$$

Thus $f \circ g$ is the function

$$\begin{aligned}f \circ g : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2\\ (x, y) &\longmapsto (-x, -y).\end{aligned}$$

(In this case, $f \circ g = g \circ f$.)

Solution to Exercise A37

The rule of $g \circ f$ is

$$\begin{aligned}(g \circ f)(x) = g(f(x)) &= g(3x + 1)\\ &= \frac{3}{(3x + 1) + 2}\\ &= \frac{1}{x + 1}.\end{aligned}$$

The domain of $g \circ f$ is

$$\{x \in [-1, 1] : f(x) \in \mathbb{R} - \{-2\}\}.$$

If $x \in [-1, 1]$, then $f(x) \in \mathbb{R} - \{-2\}$ unless $f(x) = -2$. Now $f(x) = -2$ when

$$3x + 1 = -2,$$

that is, when

$$x = -1.$$

So the domain of $g \circ f$ is

$$[-1, 1] - \{-1\} = (-1, 1].$$

Thus $g \circ f$ is the function

$$\begin{aligned}g \circ f : (-1, 1] &\longrightarrow \mathbb{R}\\ x &\longmapsto \frac{1}{x + 1}.\end{aligned}$$

Solution to Exercise A38

The domain of f is $\mathbb{R}$, and for each $x \in \mathbb{R}$ we have

$$g(f(x)) = g\,(5x - 3) = \frac{(5x - 3) + 3}{5} = x;$$

that is, $g \circ f = i_{\mathbb{R}}$.

The domain of g is also $\mathbb{R}$, and for each $y \in \mathbb{R}$ we have

$$f(g(y)) = f\left(\frac{y + 3}{5}\right) = 5\left(\frac{y + 3}{5}\right) - 3 = y;$$

that is, $f \circ g = i_{\mathbb{R}}$.

Since $g \circ f = i_{\mathbb{R}}$ and $f \circ g = i_{\mathbb{R}}$, it follows that g is the inverse function of f.

Solution to Exercise A39

This is a translation of the plane that shifts each point to the left by 1 unit and up by 3 units, so we expect its inverse to shift the plane to the right by 1 unit and down by 3 units.

Let

$$\begin{aligned}g : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2\\ (x, y) &\longmapsto (x + 1, y - 3).\end{aligned}$$

The domain of f is $\mathbb{R}^2$, and for each $(x, y) \in \mathbb{R}^2$ we have

$$\begin{aligned}g(f(x, y)) &= g(x - 1, y + 3)\\ &= (x - 1 + 1, y + 3 - 3)\\ &= (x, y);\end{aligned}$$

that is, $g \circ f = i_{\mathbb{R}^2}$.

The domain of g is also $\mathbb{R}^2$, and for each $(x, y) \in \mathbb{R}^2$ we have

$$\begin{aligned}f(g(x, y)) &= f(x + 1, y - 3)\\ &= (x + 1 - 1, y - 3 + 3)\\ &= (x, y);\end{aligned}$$

that is, $f \circ g = i_{\mathbb{R}^2}$.

Since $g \circ f = i_{\mathbb{R}^2}$ and $f \circ g = i_{\mathbb{R}^2}$, it follows that g is the inverse function of f.

Solution to Exercise A40

The vector $\mathbf{d}$ is in the same direction as $\mathbf{a}$, but none of the other vectors is; also, the magnitude of $\mathbf{d}$ is two-thirds that of $\mathbf{a}$. Hence

$$\mathbf{d} = \tfrac{2}{3}\mathbf{a} \quad \text{and} \quad \mathbf{a} = \tfrac{3}{2}\mathbf{d}.$$

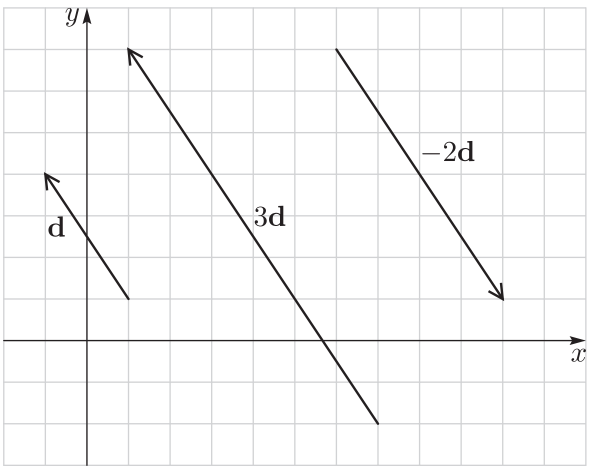
Next, $\mathbf{e}$ is parallel to $\mathbf{b}$ but in the opposite direction; none of the others is parallel to these two vectors. Also, the magnitude of $\mathbf{e}$ is three times that of $\mathbf{b}$. Hence

$$\mathbf{e} = -3\mathbf{b} \quad \text{and} \quad \mathbf{b} = -\tfrac{1}{3}\mathbf{e}.$$

Finally, $\mathbf{c}$ and $\mathbf{f}$ are not multiples of any of the other vectors.

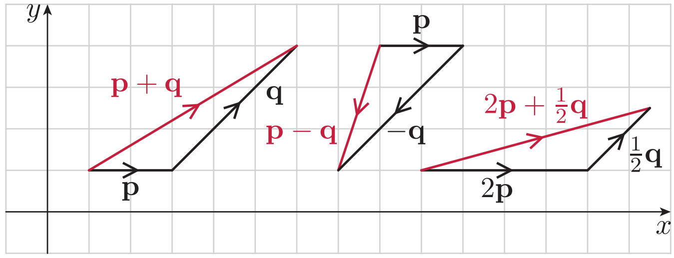
Solution to Exercise A41

The vector $3\mathbf{d}$ is in the same direction as $\mathbf{d}$, but its magnitude is three times that of $\mathbf{d}$; the vector $-2\mathbf{d}$ is in the opposite direction to that of $\mathbf{d}$, and its magnitude is twice that of $\mathbf{d}$.

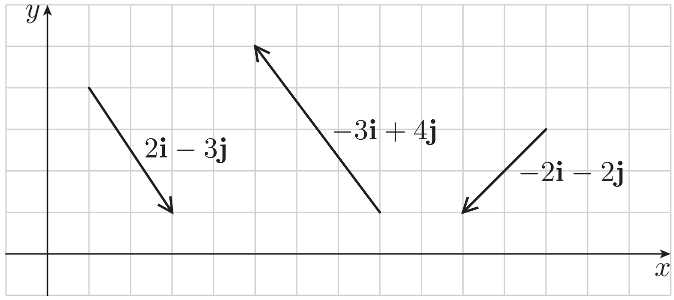


Solution to Exercise A42

We use the rule for forming a scalar multiple of a vector, and the Triangle Law for the addition of vectors.



Solution to Exercise A43



Solution to Exercise A44

(a) Here $\mathbf{p} = (3, -1)$ and $\mathbf{q} = (-1, -2)$, so

$$\begin{aligned} \mathbf{p} + \mathbf{q} &= (3 + (-1), -1 + (-2)) = (2, -3), \\ -\mathbf{q} &= (1, 2), \\ \mathbf{p} - \mathbf{q} &= (3 - (-1), -1 - (-2)) = (4, 1). \end{aligned}$$

(b) Here $\mathbf{p} = -\mathbf{i} - 2\mathbf{j}$ and $\mathbf{q} = 2\mathbf{i} - \mathbf{j}$, so

$$\begin{aligned} \mathbf{p} + \mathbf{q} &= (-1 + 2)\mathbf{i} + (-2 + (-1))\mathbf{j} = \mathbf{i} - 3\mathbf{j}, \\ -\mathbf{q} &= -2\mathbf{i} + \mathbf{j}, \\ \mathbf{p} - \mathbf{q} &= (-1 - 2)\mathbf{i} + (-2 - (-1))\mathbf{j} = -3\mathbf{i} - \mathbf{j}. \end{aligned}$$

(c) Here $\mathbf{p} = -\mathbf{i} + 2\mathbf{k}$ and $\mathbf{q} = \mathbf{i} - 2\mathbf{j} - \mathbf{k}$, so

$$\begin{aligned} \mathbf{p} + \mathbf{q} &= (-1 + 1)\mathbf{i} - 2\mathbf{j} + (2 - 1)\mathbf{k} = -2\mathbf{j} + \mathbf{k}, \\ -\mathbf{q} &= -\mathbf{i} + 2\mathbf{j} + \mathbf{k}, \\ \mathbf{p} - \mathbf{q} &= (-1 - 1)\mathbf{i} - (-2\mathbf{j}) + (2 - (-1))\mathbf{k}, \\ &= -2\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}. \end{aligned}$$

Solution to Exercise A45

(a) Since $\mathbf{p} = (3, -1)$ and $\mathbf{q} = (-1, -2)$,

$$\begin{aligned} 2\mathbf{p} &= (6, -2), \\ 3\mathbf{q} &= (-3, -6), \\ 2\mathbf{p} - 3\mathbf{q} &= (9, 4). \end{aligned}$$

The magnitude of $\mathbf{q}$ is

$$|\mathbf{q}| = \sqrt{(-1)^2 + (-2)^2} = \sqrt{5}.$$

(b) Since $\mathbf{p} = -\mathbf{i} + 2\mathbf{k}$ and $\mathbf{q} = \mathbf{i} - 2\mathbf{j} - \mathbf{k}$,

$$\begin{aligned} 2\mathbf{p} &= -2\mathbf{i} + 4\mathbf{k}, \\ 3\mathbf{q} &= 3\mathbf{i} - 6\mathbf{j} - 3\mathbf{k}, \\ 2\mathbf{p} - 3\mathbf{q} &= -5\mathbf{i} + 6\mathbf{j} + 7\mathbf{k}. \end{aligned}$$

The magnitude of $\mathbf{q}$ is

$$|\mathbf{q}| = \sqrt{(1)^2 + (-2)^2 + (-1)^2} = \sqrt{6}.$$

Solution to Exercise A46

(a) When $\mathbf{v} = (2, -3)$, the magnitude of $\mathbf{v}$ is

$$|\mathbf{v} = \sqrt{2^2 + (-3)^2} = \sqrt{4+9} = \sqrt{13},$$

so

$$\widehat{\mathbf{v}} = \frac{1}{\sqrt{13}}(2, -3) = \left(\frac{2}{\sqrt{13}}, -\frac{3}{\sqrt{13}}\right).$$

(b) When $\mathbf{v} = 5\mathbf{i} + 12\mathbf{j}$, the magnitude of $\mathbf{v}$ is

$$|\mathbf{v}| = \sqrt{5^2 + 12^2} = \sqrt{25 + 144} = 13,$$

so

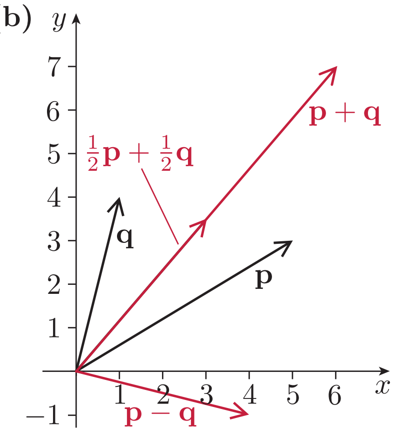
$$\widehat{\mathbf{v}} = \frac{1}{|\mathbf{v}|}\mathbf{v} = \frac{5}{13}\mathbf{i} + \frac{12}{13}\mathbf{j}.$$

Solution to Exercise A47

(a) Since $\mathbf{p} = (5, 3)$ and $\mathbf{q} = (1, 4)$,

$$\begin{aligned} \mathbf{p} - \mathbf{q} &= (4, -1), \\ \mathbf{p} + \mathbf{q} &= (6, 7), \\ \tfrac{1}{2}\mathbf{p} + \tfrac{1}{2}\mathbf{q} &= \left(\tfrac{5}{2}, \tfrac{3}{2}\right) + \left(\tfrac{1}{2}, 2\right) = \left(3, \tfrac{7}{2}\right). \end{aligned}$$

(b)



Solution to Exercise A48

(a) The vector form of the equation of l is

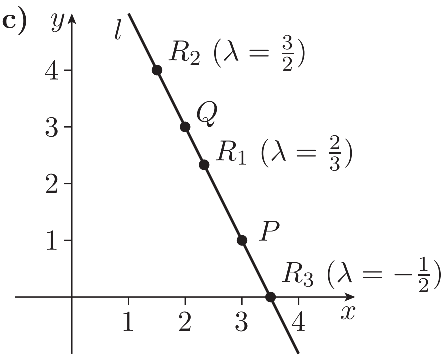
$$\begin{aligned} \mathbf{r} &= (1 - \lambda)(3, 1) + \lambda(2, 3) \\ &= (3 - \lambda, 1 + 2\lambda). \end{aligned}$$

(b) Using the formula above with $\lambda = \frac{2}{3}$, $\frac{3}{2}$ and $-\frac{1}{2}$ in turn, we obtain the following position vectors:

$$\begin{aligned} \mathbf{r}_1 &= \left(3 - \tfrac{2}{3}, 1 + \tfrac{4}{3}\right) = \left(\tfrac{7}{3}, \tfrac{7}{3}\right), \\ \mathbf{r}_2 &= \left(3 - \tfrac{3}{2}, 1 + 3\right) = \left(\tfrac{3}{2}, 4\right), \\ \mathbf{r}_3 &= \left(3 - \left(-\tfrac{1}{2}\right), 1 + (-1)\right) = \left(\tfrac{7}{2}, 0\right). \end{aligned}$$

Thus the three points on the line are the points R_1, R_2 and R_3, with coordinates $\left(\frac{7}{3}, \frac{7}{3}\right)$, $\left(\frac{3}{2}, 4\right)$ and $\left(\frac{7}{2}, 0\right)$, respectively.

(c)



Solution to Exercise A49

(a) The vector form of the equation of l is

$$\mathbf{r} = (3 - \lambda, 1 + 2\lambda).$$

Hence at the point $(4, -1)$ on l, we have

$$(4, -1) = (3 - \lambda, 1 + 2\lambda).$$

Equating the corresponding components gives

$$4 = 3 - \lambda \quad \text{and} \quad -1 = 1 + 2\lambda.$$

The first equation gives $\lambda = -1$, and this value of λ also satisfies the other equation. Hence the value of λ corresponding to the point $(4, -1)$ in the vector form of the equation of l is $\lambda = -1$.

(b) The point $\left(\frac{1}{2}, \frac{1}{2}\right)$ lies on l if and only if there is some real number λ for which

$$\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = (3 - \lambda, 1 + 2\lambda).$$

Equating corresponding components gives

$$3 - \lambda = \tfrac{1}{2} \quad \text{and} \quad 1 + 2\lambda = \tfrac{1}{2}.$$

The first of these equations has solution $\lambda = \frac{5}{2}$, and the second has solution $\lambda = -\frac{1}{4}$.

It follows that there is no real number λ that satisfies the vector form of the equation of l, when $\mathbf{r} = \left(\frac{1}{2}, \frac{1}{2}\right)$, so the point $\left(\frac{1}{2}, \frac{1}{2}\right)$ does not lie on l.

Solution to Exercise A50

(a) The vector form of the equation of the line l is

$$\begin{aligned}\mathbf{r} &= (1 - \lambda)(2, 1, 0) + \lambda(1, 0, -1)\\ &= (2 - \lambda,\ 1 - \lambda,\ -\lambda).\end{aligned}$$

(b) Using the formula above with $\lambda = \frac{1}{2}$ and -1, we obtain the following position vectors:

$$\begin{aligned}\mathbf{r}_1 &= \left(2 - \tfrac{1}{2},\ 1 - \tfrac{1}{2},\ -\tfrac{1}{2}\right)\\ &= \left(\tfrac{3}{2}, \tfrac{1}{2}, -\tfrac{1}{2}\right),\\ \mathbf{r}_2 &= (2 - (-1),\ 1 - (-1),\ -(-1))\\ &= (3, 2, 1).\end{aligned}$$

Thus the two points have coordinates $\left(\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}\right)$ and $(3, 2, 1)$.

Solution to Exercise A51

We use the formula for the scalar product of vectors in component form.

(a) $$\begin{aligned}(2, 3) \cdot \left(\tfrac{5}{2}, -4\right) &= 2 \times \tfrac{5}{2} + 3 \times (-4)\\ &= 5 - 12 = -7\end{aligned}$$

(b) $$\begin{aligned}(1, 4) \cdot \left(2, -\tfrac{1}{2}\right) &= 1 \times 2 + 4 \times \left(-\tfrac{1}{2}\right)\\ &= 2 - 2 = 0\end{aligned}$$

(c) $$\begin{aligned}&(-2\mathbf{i} + \mathbf{j}) \cdot (3\mathbf{i} - 2\mathbf{j})\\ &\quad = (-2) \times 3 + 1 \times (-2)\\ &\quad = -6 - 2 = -8\end{aligned}$$

(d) $$\begin{aligned}&(1, -1, -2) \cdot (3, -2, -5)\\ &\quad = 1 \times 3 + (-1) \times (-2) + (-2) \times (-5)\\ &\quad = 3 + 2 + 10 = 15\end{aligned}$$

Solution to Exercise A52

In each case we let $\mathbf{u}$ denote the first vector of the pair, $\mathbf{v}$ the second vector, and θ the angle between the two vectors.

(a) Here

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} &= (1, 4) \cdot (5, 2) = 5 + 8 = 13,\\ |\mathbf{u}| &= \sqrt{1^2 + 4^2} = \sqrt{1 + 16} = \sqrt{17},\\ |\mathbf{v}| &= \sqrt{5^2 + 2^2} = \sqrt{25 + 4} = \sqrt{29}.\end{aligned}$$

Hence

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \frac{13}{\sqrt{17}\sqrt{29}} = \frac{13}{\sqrt{493}},$$

so

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{13}{\sqrt{493}}\right)\\ &= 0.95 \text{ radians} \quad (\text{to 2 d.p.}).\end{aligned}$$

(b) Here

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} &= (-2, 2) \cdot (1, -1) = -2 - 2 = -4,\\ |\mathbf{u}| &= \sqrt{(-2)^2 + 2^2} = \sqrt{4 + 4} = \sqrt{8} = 2\sqrt{2},\\ |\mathbf{v}| &= \sqrt{1^2 + (-1)^2} = \sqrt{1 + 1} = \sqrt{2}.\end{aligned}$$

Hence

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \frac{-4}{2\sqrt{2}\sqrt{2}} = -1,$$

so

$$\theta = \cos^{-1}(-1) = \pi \text{ radians}.$$

You might have expected this result, because $\mathbf{u}$ and $\mathbf{v}$ point in opposite directions (in fact, $\mathbf{u} = -2\mathbf{v}$).

(c) Here

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} &= (9\mathbf{i} - 2\mathbf{j}) \cdot (\mathbf{i} + 2\mathbf{j})\\ &= 9 \times 1 + (-2) \times 2\\ &= 9 - 4 = 5,\\ |\mathbf{u}| &= \sqrt{9^2 + (-2)^2} = \sqrt{81 + 4} = \sqrt{85},\\ |\mathbf{v}| &= \sqrt{1^2 + 2^2} = \sqrt{5}.\end{aligned}$$

Hence

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \frac{5}{\sqrt{85}\sqrt{5}} = \frac{1}{\sqrt{17}},$$

so

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{1}{\sqrt{17}}\right)\\ &= 1.33 \text{ radians} \quad (\text{to 2 d.p.}).\end{aligned}$$

Solution to Exercise A53

In each case we let $\mathbf{u}$ denote the first vector of the pair, $\mathbf{v}$ the second vector, and θ the angle between the two vectors.

(a) Here

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} &= (3,4,5)\cdot(1,0,-1)\\ &= 3\times 1 + 4\times 0 + 5\times(-1) = -2,\end{aligned}$$

$$|\mathbf{u}| = \sqrt{3^2+4^2+5^2} = \sqrt{50} = 5\sqrt{2},$$

$$|\mathbf{v}| = \sqrt{1^2+0^2+(-1)^2} = \sqrt{2}.$$

Hence

$$\begin{aligned}\cos\theta &= \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|}\\ &= \frac{-2}{5\sqrt{2}\sqrt{2}} = -\frac{1}{5},\end{aligned}$$

so

$$\begin{aligned}\theta &= \cos^{-1}\left(-\frac{1}{5}\right)\\ &= 1.77 \text{ radians} \quad (\text{to 2 d.p.}).\end{aligned}$$

(b) Here

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} &= (2\mathbf{j}-3\mathbf{k})\cdot(-\mathbf{i}-\mathbf{j}-2\mathbf{k})\\ &= 0\times(-1) + 2\times(-1) + (-3)\times(-2)\\ &= -2+6 = 4,\end{aligned}$$

$$\begin{aligned}|\mathbf{u}| &= \sqrt{0^2+2^2+(-3)^2}\\ &= \sqrt{4+9} = \sqrt{13}\end{aligned}$$

$$\begin{aligned}|\mathbf{v}| &= \sqrt{(-1)^2+(-1)^2+(-2)^2}\\ &= \sqrt{1+1+4} = \sqrt{6}.\end{aligned}$$

Hence

$$\begin{aligned}\cos\theta &= \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|}\\ &= \frac{4}{\sqrt{13}\sqrt{6}} = \frac{4}{\sqrt{78}},\end{aligned}$$

so

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{4}{\sqrt{78}}\right)\\ &= 1.10 \text{ radians} \quad (\text{to 2 d.p.}).\end{aligned}$$

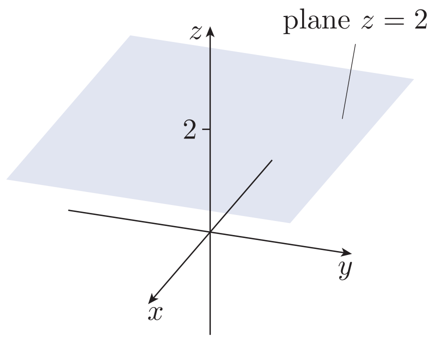
Solution to Exercise A54

Points (x,y,z) that lie in the (y,z)-plane all have $x=0$; so $x=0$ is the equation of this plane.

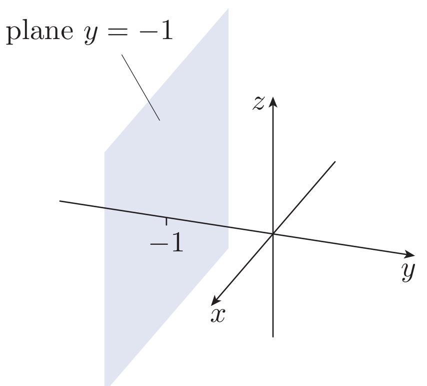
Similarly, points (x,y,z) that lie in the (x,z)-plane all have $y=0$; so $y=0$ is the equation of this plane.

Solution to Exercise A55

(a) This plane is parallel to the (x,y)-plane and passes through the point $(0,0,2)$.



(b) This plane is parallel to the (x,z)-plane and passes through the point $(0,-1,0)$.



Solution to Exercise A56

We use the formula

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{p} \cdot \mathbf{n}$$

for the equation of a plane, where $\mathbf{x} = (x, y, z)$, $\mathbf{n}$ is a normal to the plane and $\mathbf{p}$ is a point in the plane.

(a) Here $\mathbf{n} = (2, 3, 1)$ and $\mathbf{p} = (1, 0, 2)$, so the equation of the plane is

$$(x, y, z) \cdot (2, 3, 1) = (1, 0, 2) \cdot (2, 3, 1).$$

This can be expressed in the form

$$2x + 3y + z = 1 \times 2 + 0 \times 3 + 2 \times 1,$$

that is,

$$2x + 3y + z = 4.$$

(b) Here $\mathbf{n} = (4, -2, 1)$ and $\mathbf{p} = (-1, 1, 5)$, so the equation of the plane is

$$(x, y, z) \cdot (4, -2, 1) = (-1, 1, 5) \cdot (4, -2, 1).$$

This can be expressed in the form

$$4x - 2y + z = (-1) \times 4 + 1 \times (-2) + 5 \times 1,$$

that is,

$$4x - 2y + z = -1.$$